

Reprinted from Psychological Review Monograph No. 69



Accuracy of the Voice in Simple Pitch Singing

BY

WALTER R. MILES

3781
I65M

UNCLASSED

Dissertation



Reprinted from Psychological Review Monograph No. 69

Accuracy of the Voice in Simple Pitch Singing

BY

WALTER R. MILES

A thesis submitted to the Department of Philosophy and Psychology of the Graduate College in the State University of Iowa, in partial fulfillment of the Requirement for the degree of Doctor of Philosophy.

ACCURACY OF THE VOICE IN SIMPLE PITCH SINGING

BY

WALTER R. MILES

Contents

Historical

The tonoscope

Experiments Series I: accuracy and the voice range

Experiments Series II: intensity of standard

Experiments Series III: volume of the voice

Experiments Series IV: timbre of standard tones

Experiments Series V: vowel quality and accuracy

Experiments Series VI: accuracy in singing

Apparatus and method: standards, observers, the charge, the test.

Justification of procedure: voice level of test, forks for standards,
many standards vs. one standard, sounding the two tones, order
of standards, time intervals, other factors

Tables of data

Comparison of the abilities of men and women

The mean variation

The constant error

The group constant error

The best cases

Correlation of singing with pitch discrimination

The test of 1910

Recommendations

Summary of conclusions

References

The experiments here reported deal with two phases of simple pitch singing: (1) the ability of the voice to reproduce the pitch of a tone, and (2) the ability to make faint shadings in pitch, sharp or flat. The aim has been to formulate, if possible, a standard test for the measurement of each, to establish norms, and to investigate some of the underlying psychological factors.¹

¹ The extensive measurements made would have been impossible were it not for the previous labor of Professor Seashore in perfecting a recording apparatus, the Tonoscope. Dr. Seashore has furthermore proved himself an unfailing source of inspiration and suggestion throughout the experimentation. The author is also under heavy obligations to Assistant Professor Mabel C. Williams, Dr. Thomas F. Vance, Messrs. Bruene and Malmberg, and the many observers for their kind and prolonged assistance.

HISTORICAL

The first investigator to employ the experimental method in attacking the problem of the accuracy of the voice in singing pitch was Klünder (11) 1872.² He used a manometric flame with two connected speaking tubes, an organ tone sounding in one while the observer sang simultaneously in the other. The difference in vibration number between the standard and sung tones was determined by counting waves. The average $\pm$ errors found on three tones, 128, 192, 256, v.d. are 0.761, 0.434, and 0.257 per cent. (of standards) respectively. The difference between 0.761 and 0.257 was thought to be due to the vocal cords and not to hearing.

Klünder was not satisfied with his method or his results and continued working on the problem, publishing a second time in 1879 (12). Again he used organ tones as standards and had his observers sing simultaneously with them, either in unison or in specified interval. The recording was done on smoked paper by means of two phonautographs. The two records were compared directly, that for the organ tone being used as a standard, and deviation in the pitch of the voice from that of the standard was computed in terms of .25 v.d. That Klünder was primarily interested in the physiological side of the problem is indicated by the questions which he set himself:

(1) Does our ear control the voice or is it controlled by the feeling of tension in the larynx? (2) How firmly does the voice attack tones? (3) Are the fluctuations of the voice such that give proof of control by the ear? (4) How many stress degrees of muscular tetanus are we justified in accepting through the performance of the muscles of the larynx?

Klünder found that for the pitches 96, 128, 192, 256, v.d., respectively, he himself as observer made the following $\pm$ errors: .32 v.d., .47 v.d., .62 v.d., and .59 v.d. This however was somewhat better than any of his other observers could do.

From this Klünder concludes that the voice is very accurate in reproduction of pitch and he answers his questions in substance as follows:

² Previous to this time Scott (17) and Blake (3) had developed phonautographs for registering voice curves.

(1) The vocal cords are held in labial tension by muscular tetanus. (2) The musculature allows from 40 to 170 different tensions in the tetanus. (3) The regulation of the pitch of the voice takes place directly through the sensation of tension in the larynx.

Seashore (19) in 1910 published in a very condensed form the results of experiments carried on in 1905 by himself and E. A. Jenner. Previous to that time, however, much work had been done in perfecting a registering apparatus, the tonoscope, which is fully described by Professor Seashore in the foregoing article in this volume of the *Studies*. Some preliminary experimenting also was done in 1901-'02 with the help of Edward Bechly, the results of which have never been published. Seashore and Jenner in their work sought to answer two questions: (1) Can we facilitate development of control in the pitch of the voice by using an aid to the ear in training? (2) May the ordinary limits of accuracy be exceeded by training with such an aid? In attacking these problems they used three measurements: (1) accuracy in reproducing a given tone, (2) accuracy in singing a required interval, and (3) the least producible change in the pitch of the voice. The standard or fundamental tone was 100 v.d., produced by a large tuning-fork; the intervals were the major third, the fifth, and the octave above this. The least producible change was determined for each of these four tones (1) in the least producible sharp and (2) in the least producible flat from the note as actually sung. Each period of practice consisted of one hundred and sixty trials, which took about forty-five minutes. The tests continued twelve days, approximately successive. During the first five days the singer depended entirely on the ear as in ordinary singing: then followed five days of singing with aid, *i.e.*, the observer was informed of the result of each trial immediately after it was made. The records of the eleventh day were taken without aid, while on the twelfth day aid was again given. Six men acted as observers. The conclusions of this investigation are quoted as follows:

"(1). The aid enhances the ability to strike a tone which has been heard. The superiority of the aided series over the unaided amounts to 42 per cent. (2) The aid enhances the ability to sing an interval. The superiority of the aided series over the unaided amounts to 50 per cent. for the major third, 50 per cent. for the fifth, and 60 per cent. for the octave. (3) The voluntary control

of the pitch of the voice is improved by the aid. The average superiority of the aided series over the unaided for all intervals amounts to 26 per cent. (4) There is probably some transfer of gain from the aided training to following unaided singing. (5) There is no evidence of transfer of the gain in the accuracy of the memory image. This is undoubtedly due to the fact we have here to do with memory rather than discrimination and the acquisition of accurate memory images is a slow process—too slow in this short series. (6) The gain in the discriminative control of pitch of the voice is fully transferred. (7) Improvements in the ability to sing a tone or an interval, and the ability to produce a minimal change, are very much more pronounced and more rapid in the aided than in the unaided series. (8) The second question is not answered absolutely by our records, but it seems probable (a) from the radical and immediate improvement of the aided series over the unaided, and, (b) from the introspection showing a tone which without the instrument seemed entirely satisfactory to the ear could be corrected by the ear after the error had been pointed out by the instrument, that a higher degree of accuracy of pitch in singing may be attained by aiding the ear in the training than would be possible to attain without such aid. No matter how keen the ear of a trained musician, it can be shown in a single test that his ear has been “too generous”—too easily satisfied, for when the error is pointed out objectively he can recognize it. We thus find cumulative evidence to show that the singer can not reach the physiological limit of accuracy by the ordinary methods of voice culture, because he has no objective criterion by which he can check up the accuracy of his ear. (9) The major third, the fifth, and the octave are approximately equally difficult intervals to sing. If we express the average error in relative fractions of a tone ($1/25$ of a tone) instead of in vibrations, the ratio is 1.4, 1.5, and 1.4, for the three intervals named above. The average error expressed in terms of vibrations shows that the difficulty of a natural interval varies approximately with the magnitude of the interval. (10) The minimal change is a relatively constant fraction of a tone within the octave. This is true for both the aided and the unaided series. If we reduce the records from vibrations to twenty-fifths of a tone, the minimal change is 3.1, 3.1, 3.6, 3.3, for the fundamental, the major third, the fifth, and the octave respectively. This is surprising,

because within this part of the tonal range the pitch discrimination is normally measured by a constant vibration frequency instead of by a constant part of a tone."

Cameron (4) 1907, varied somewhat the conditions of the experiment as performed by Klünder. In the first series the subject was asked to sing any tone of medium pitch, a second tone of low pitch, and a third of high pitch, and to sustain the pitch selected in each case as uniformly as possible throughout the singing. The second series was like the first except that each tone was interrupted by the insertion of short pauses of .3 second duration. In a third series, somewhat longer than those previously mentioned, the ability of one observer to imitate organ tones in the range 94 v.d. to 303 v.d. was tested. The tones were reproduced in sequence and, in chance order, partly simultaneously with the standards and partly by singing the tones immediately after the organ had ceased sounding. In a fourth series various distracting tones, (1) harmonious or inharmonious with the standard tone; (2) of greater or less interval from the standard; and (3) higher or lower than the standard, were introduced either at the beginning or just preceding the beginning of the reproduction by the observer. The more important results of the study are here summarized.

"(1) In the singing of a tone a sudden marked rise in pitch usually occurs near the beginning of the tone. This rise in pitch is so general as to seem to indicate a universal tendency. (2) No tone is sung entirely uniformly. It oscillates in pitch from period to period throughout its length in a somewhat irregular rhythmical fashion. (3) Very marked differences exist in different individuals with regard to their ability to imitate a standard tone. The subjects tested varied in degrees of accuracy in imitation of standard tones of different pitch from a small fraction of 1 per cent. to 13 per cent. of error. (4) There is manifest throughout a tendency to sing a tone higher than it should be sung. Thus the end of a tone is usually higher than the beginning and the sung tone (as a whole) is almost invariably higher than the standard tone. (5) Distractions when causing disturbances may affect the whole of the sung tone or only the beginning of the tone. In either case the effect of the distraction may be to cause the sung tone to vary from the standard (a) in the direction of the distracting tone; or (b) in the opposite direction from the distracting tone. (6) Sung tones vary-

ing from the standard under the effect or distraction are usually harmonious with the distracting tone. When the distracting tone is inharmonious with the standard tone, distraction is more likely to occur than when the two tones form a harmony. (7) A person may more or less closely imitate a tone which he has heard when his attention was engrossed in singing another tone of a standard pitch."

An important contribution to the general problem of control of the pitch of the voice in singing was made by *Berlage* (2) in 1910. During the summer of 1907 Berlage carried on a series of experiments in which definite time intervals were inserted between the breaking off of the standard tone and the beginning of the reproduction by the observer.³ These intervals were of the following values stated in seconds: 1, 2, 3, 4, 5, 7, 10, 15, 20, 25, and 30. The tones were all sounded as "a" ('a' in 'ah'). This series is an amplification of the methods of Klünder and Cameron, and was undertaken for the purpose of finding the time interval most favorable for the imitation of tones, which when found became one of the conditions of further experimentation.

In the winter of 1907-'08 Berlage's general problem was to determine the influence of articulation and hearing in the vocal reproduction of tones. In this series (second) as in the third series by Berlage the standard tones to be imitated are voice tones. The variation of conditions consisted in having the standard tones sung part of the time by the observer and part of the time by the experimenter thus showing the immediate influence of hearing and of loud articulation in tone-reproduction. It seemed desirable to determine to what extent the influence of articulation is due to the larynx, and to the mouth cavity. For this purpose, in a third series of experiments, all the standard tones were sung by the observers, the vowel quality being varied under control. The standard and reproduction were sung, sometimes to the same vowel as "i", "i", or "u", "u", and at other times to different vowels as to "i", and "u" or "a" and "u". The chief conclusions reached from Berlage's experiments are the following:

(1) "Accuracy in the reproduction of a "strange" voice tone decreases rather regularly with the increasing time interval of from 1 to 30 seconds. Accuracy is greatest with an interval of from

³ Berlage designates these tones as 'foretone' and 'aftertone'. 'Standard' and 'reproduction' are used throughout this study.

1 to 2 seconds. The values found here, for the variable average error, in the case of the observers amounted to only .5 v.d. and .6 v.d. (2) Observers reproduced their own voice tones more accurately than those of another (time interval 3 seconds). (3) The increase of precision shows itself chiefly in a decrease of the constant error. In the reproduction of outside standards and especially when they are near the boundaries of the voice range there is a tendency toward a constant error near the middle of the voice range. (4) In the reproduction of one's own tones vowel change works a disadvantage upon precision. With the standard tone sung as "u" and the reproduction as "i" there is a tendency for the latter to be lower, and vice versa when the vowels are changed. (5) In the reproduction of an outside standard the variable average error expressed in vibration frequency becomes larger with rising pitch, while if expressed in per cent. of vibration frequency it diminishes. (6) In the reproduction of one's own tones the variable average error expressed in vibration frequency remains rather constant with rising pitch. (7) The amount of departure of the individual tone sections (measured off in .1 second periods) from the general average of the reproduction shows no tendency, in the variations carried out in these experiments, to change according to the ordinal number of the tone sections in the course. (8) Only in the first .1 second is the reproduction regularly lower than the rest of the tone course. (9) Reproductions after the time intervals of from 3 to 10 seconds, in the case of two observers, show a sudden raising or lowering of the tone after the tone has progressed some .4 to 1.2 seconds. (10) The average departure of individual tone sections from the average for the tone is greatest in the reproduction of one's own tones. (11) The total amount of departure, expressed in vibration frequency grows with rising pitch so that—not considering rather marked irregularities with the individual observers—the amount of variation expressed in per cent. of a tone remains about constant."

The latest published study of this general problem to come to our attention is that of *Sokolówsky* (22) 1911. His apparatus consisted of a combination of the Einthoven string-galvanometer and the Weiss phonoscope. The organ tones, which were used for standards, acted on the string-galvanometer and the sung tones on the phonoscope. Both tones were registered in a convenient way for comparison by means of the Blix-Sandström photokymograph.

Sokolowsky secured the coöperation of seven professional opera singers, three men and four women. The observers were allowed to choose the vowel to which they sang the tones. The musical "a" was chosen most frequently. There were three short series of experiments: (1) singing a given tone simultaneously with the sounding of the tone by the organ (unison); (2) allowing a time interval between the organ tone and its reproduction. (The intervals used were 30, 60, and 120 seconds, during which the observers were instructed not to hum or sing to themselves); and (3) singing a specified interval from a simultaneously sounding organ tone. The musical intervals selected were the third, fourth, fifth, sixth and octave.

The results from these three series of experiments may be summarized as follows: (1) Curves for 8 tones were secured in series I. The average pitch was 251 v.d. (range 165 to 296 v.d.), the average error was ± 0.44 per cent. The average pitches for men and women respectively were 197 and 286 v.d., with average $\pm$ errors of 0.51 and 0.40 per cent. (2) The introduction of a time interval increases the average error to ± 0.99 per cent. as compared with ± 0.44 of the previous series. Errors are usually larger with an interval of 60 seconds than with 30 seconds. (3) The average error in series III is ± 1.51 per cent. The largest errors, average ± 3.28 per cent. are on the fifth, while the smallest, average ± 0.78 per cent. are on the third. (4) Of the entire number of tones counted (46) 36 are sung flat and 10 sharp. The errors on the side of sharpening are divided among three women and one man; those on the side of flattening between three men and three women.

Guttmann (6) 1912, in his consideration of the psychophysics of singing gives some attention to the problem of accuracy in reproducing pitch and states that recently he has been engaged in an extensive research in this field. The results are to be published shortly in one of the psychological journals, but in a preliminary way he says that they agree in general with those secured by Klünder and Sokolowsky, but he thinks that the results of the latter (unison curves) are "too good".

Other investigators, among them *Hensen* (10) and more recently *Marbe* (14), *Grützner* (5) and *Scripture* (18) have developed methods for recording the pitch of the voice, but these seem not to have been used in gathering data on our problem.

THE TONOSCOPE

In the investigations of Klünder, Cameron and Berlage the vibration frequency of the tones was recorded in tracings on smoked paper. Sokolowsky photographed his records; after these had been rendered permanent the waves were counted and the pitch determined by comparison with a time or standard line. This method, commonly known as "graphic recording" has been used with various refinements by many investigators in the field of phonetics. While reliable, it is at best indirect and very laborious.

Seashore and Jenner in their research made use of an early model (20) of the tonoscope. This instrument as lately improved was used by the author in the present experiments.⁴ It has several advantages which recommended it as an instrument for the measurement of the pitch of tones. In the first place readings are made quickly and directly. The instant a tone is sounded the vibration frequency is indicated by a row of dots. The experimenter has simply to note the number of this row and to record it. He is, therefore, enabled to secure a large number of observations in a relatively short time. It is not difficult to take two hundred records in thirty minutes. In the second place the experimenter has the advantage of knowing how the test is progressing. If a preliminary practice series is desired to acquaint the observer with some procedure we have in the direct readings from the tonoscope an index to the observer's understanding of the test. The observer must be kept actively trying throughout the experiment. In psychological tests, such as the imitation of tones by singing, there is so much repetition in the program for the observer that his attention easily wanders. Large and unnatural errors are therefore likely to be found in the records. Here the tonoscope as a recording instrument has an advantage over other methods as it provides for detecting these errors as soon as they occur. The experimenter as he takes each reading notes roughly the attack, the steadiness, and the degrees of success with which the reproduction approaches the standard. He thus easily becomes acquainted with the unusual range of variability and forms an estimate of the observer's power to control his voice. When a

⁴The instrument is fully described in the preceding article in this volume of *Studies*, "The Tonoscope", by Professor Seashore. A reading of that article is essential for an understanding of the present report.

tone of unusual divergence is given he therefore immediately recognizes it and can take cognizance of it, asking for introspection or for a new trial, and all with scarcely any loss of time. He may thus check up and to some extent control the observer,—keep him at his best. Furthermore the possibility of encouraging the observer or even of giving him full information regarding the success or failure of each trial is in itself a most important asset.

The tonoscope has been criticised as giving only an approximate result, because the pitch of the singing voice is not uniform and it is therefore necessary in reading the instrument to select the predominating pitch. This criticism stands or falls according to the needs of the problem to be attacked. If one were studying the oscillations of the voice, or the variations of the individual sections of a tone, as for example the difference in pitch between the first tenth and the fifth tenth of a second of a tone, it would be better to use a graphic method. But even in such problems as these the tonoscope is not without its possibilities. The characteristics of tonal attack in singing are easily discernible in the configurations on the screen. With many of the problems which lie in our field there is no need for so detailed a record. The predominant or modal pitch of a tone of from one to two seconds in length is all that is needed for much of the work in the psychology of pitch singing. The tonoscope can of course meet this condition admirably, as it is this modal pitch which stands out clear and distinct, forcing itself upon the attention of the experimenter.

Tonoscope reading test.—The method of reading the tonoscope, and the various sources of error having been fully treated by Professor Seashore in the accompanying article, there is no need to repeat them here.

In order to determine the degree of accuracy in the reading of the tonoscope the following experiment was performed. A set of ten large, movable-disc, tuning forks ranging from 128 to 131 v.d. was so tuned that no two forks had a pitch difference of over 3 v.d. and in the great majority of cases the differences were much smaller. A revolving shutter, rotated by the tonoscope shaft, was so arranged as to expose the mouth of a resonator connected with the sensitive light for the following time intervals: .25, .50, .75 and 1.00 second. In this way a tone sounded before the shutter was registered by the tonoscope for just the period during which the

resonator was exposed.⁵ The presentation of the tones and the recording of the observations were in charge of two helpers. The experimenter did nothing but watch the moving screen and call out the readings. He had no way of knowing the real reading in any case. Five trials were given on each fork with each exposure interval. The order of the forks was determined approximately by chance. There was an interval of about five seconds between tones.

After the fifty trials with the .75 second exposures were finished, the pitch of each fork was carefully determined with the tonoscope, counts being made by the stop-watch during periods of from 6 to 15 seconds. These records formed the basis from which to compute the errors in the first test. The assistant then changed the pitch of all the forks and the above procedure was repeated with a .50 second exposure. Again the forks were changed and the same procedure was followed for the 1.00 second and the .25 second exposures in turn. Thus fifty records were obtained for each of the four exposure periods and the conditions were such that the reader could have no accessory clue. The record is summarized in Table I.

To test the reading ability for tones one octave higher, *i.e.* 256 v.d., where it will be recalled the tonoscope reading, and hence the errors, must be doubled, a set of seven small forks was provided. These were weighted so that no pitch difference between any two forks was greater than 3 v.d. The test was made with the exposure interval of .75 second.

In making the pitch difference between the forks come within a range of 3 v.d. we approximate the condition presented when working with voice tones that require accuracy in reading. If an observer is asked to reproduce a tone or to sing an interval the experimenter knows approximately the point on the scale where the reading should occur. He is watching this point. Should the reproduction be nearly correct and the tone fairly constant for, say .50 second, he can read according to our result (see Table I) within an error of less than $\pm .2$ v.d. If however the reproduction goes wide of the mark, for example to the extent of 6 v.d. there is no need of reading in fractions smaller than halves.

⁵ This arrangement is not ideal in that, as the tone is turned on and cut off by the disc, slightly disturbing waves are set up and show on the screen. In test No. 4 where the tone sounded for .25 seconds this was felt to be very disturbing. The real time given for the reading of the tones in all these tests was thus slightly less than that represented by the several discs.

TABLE I. *The degree of accuracy in the reading of the tonoscope*

Exposure	1.00 sec.		Ave.	error	.12	v.d.	; m.v.	.10	v.d.
"	.75	" (128 v.d.)	"	"	.15	"	; "	.12	"
"	.75	" (256 v.d.)	"	"	.19	"	; "	.17	"
"	.50	" (128 v.d.)	"	"	.18	"	; "	.11	"
"	.25	" (128 v.d.)	"	"	.65	"	; "	.27	"

EXPERIMENTS SERIES I: ACCURACY AND THE VOICE RANGE

In the first five series of experiments the purpose was to answer questions concerning some factors which must be considered in any adequate test of voice control. (1) How does accuracy of control vary with the range of the voice? (2) How does the intensity of the standard tone affect the pitch reproduction? (3) What is the relation of voice volume to voice control? (4) Are the reproductions affected by the timbre of the standard tones? (5) Do vowel changes (timbre changes) in the reproductions cause changes in the pitch of the reproduction? The sixth series represents an effort to combine into a single test the results of our previous experiments, together with those of other investigators, and to give this test to a sufficiently large group that we might be enabled to determine from the results some of the norms of voice control.⁶

Seventeen men with splendid enthusiasm gave their services as observers in the experiments of Series I. From among this number several were selected to serve as observers in Series II, III, IV, and V. The observers were all of mature age and more than half their number had had some training in the methods of experimental psychology. P, the only professional musician in the group, is a teacher of "Voice" and a thoroughly trained tenor soloist. H, a baritone of extensive special training, has for some time been the leader of a large choir. He is a soloist of ability. Ma., W, and V. Z. have all had special training in singing, and much experience in solo, quartette and glee club work. S, C. Mi., Ro., An., Wi., and V. H. all have had considerable experience in general singing but are without special training. Ri., Ab., Mc., Br., D, and Bh. very seldom sing in public but they enjoy music.

⁶ Gutzmann (7) and Sokolowsky (22) suggest some of the above problems, especially Nos. 1 and 5 as being important. These articles and suggestions however did not come to the attention of the writer until the experimentation was completed.

For Series I the standard tones were provided by a set of twenty tuning forks ranging approximately by the chromatic scale from C_1 , 64 v.d., to and including a' , 426 v.d. The first fourteen forks beginning with 64 v.d. were large and carried discs. All the tones were of good quality and their duration of tone was more than ample. Some of the forks were of different vibration frequency than that indicated by the notes of the chromatic scale; for example, the pitch of the fork that corresponded to G was 182 v.d. in place of 192 v.d. These differences were made in order to check the observers from judging and singing the various steps as musical intervals.

An independent selection of five forks was made for each observer after a preliminary determination of his voice range. These forks covered approximately one and one-half octaves in the middle of the range and were fairly distributed. In giving the test the experimenter presented the tones to the ear of the observer, who, after listening for 1.5 seconds and allowing a time interval of 1 second, reproduced the pitch of the tones as accurately as possible. Proceeding from the lowest to the highest and then in reverse order back to the lowest, each tone was given twice in succession, the test consisting of twenty trials on each standard tone.

The results of these experiments are present in Table II. O. denotes the observer; P, the pitch of the standard tone, E, average error, m.v., mean variation; and C.E., constant error. These five successive columns give the record of the respective standards for each observer. The footings in the table show the averages of the figures above stated, first in terms of vibration (absolute) and second, in terms of percentage of a tone (relative) at the respective levels. The average C. E. in the footing is the average of C. E. regardless of sign; in the second the sign is taken into account giving group tendency of the constant error, or group constant error (G. C. E.). These footings are represented graphically in Fig. 1.

Taken as a whole these records show that accuracy in the reproduction of the pitch of a tone, as measured by the average error (E) with its mean variation (m.v.) the average of the constant errors (C.E.) and the general tendency of the constant errors (G. C. E.), tends to be a constant in terms of vibration frequency. This is shown in Fig. 1 (A) by the fact that the four curves, for

the absolute variation tend to remain horizontal lines whereas the four curves for the relative variation (B) tend to fall in inverse ratio to the rise of the pitch. The slight tendency to deviate from the constant in terms of vibrations is in the direction of decrease in accuracy with rising pitch. This, in the case of the highest tone,

TABLE II. Accuracy and the voice range

O	P	E. m.v. C.E.	P	E. m.v. C.E.	P	E. m.v. C.E.	P	E. m.v. C.E.	P	E. m.v. C.E.
P.	128	1.6 1.1 -1.5	160	1.2 .6 -1.2	182	.9 .8 +.4	256	1.9 1.6 -1.6	320	1.2 1.2 +.4
H.	95	1.4 .6 +1.4	120	3.1 1.1 +3.1	160	.7 .7 -.1	213	2.9 1.4 -2.9	286	3.7 1.4 -3.7
Ma.	95	1.5 .7 -1.5	120	.8 .8 +.4	160	3.0 1.1 +2.9	240	3.7 1.0 +3.7	286	3.1 1.3 +3.0
W.	120	.9 .9 +.2	144	1.0 .8 +.4	182	1.7 1.6 -.2	256	2.5 1.5 +1.9	320	1.8 1.7 -.1
V.Z.	120	3.1 1.1 +3.1	144	1.6 .7 +1.6	182	1.7 .9 +1.4	256	1.1 1.1 +.3	341	2.7 1.4 +1.7
S.	120	2.6 .8 -2.4	144	1.8 .9 -1.5	182	1.0 .8 -.3	240	2.1 1.2 -.9	286	7.1 1.8 +7.1
C.	95	.9 1.2 +.2	144	.7 .7 -.0	182	.8 .6 -.3	256	1.3 1.9 +.8	341	1.7 1.5 -.9
Mi.	86	2.9 2.3 +4.0	120	1.1 .8 +1.0	182	1.6 1.9 -.1	256	2.7 2.7 +1.0	341	3.5 1.8 +3.4
Ro.	95	2.3 4.5 +.1	144	2.9 2.2 +2.5	182	2.1 2.0 +1.3	240	4.5 1.8 +4.3	320	2.5 2.0 +.2
An.	95	5.0 2.3 +3.7	120	1.6 1.5 +.8	160	2.5 2.8 -1.0	213	4.5 2.6 -4.1	256	4.6 3.5 -3.1
V.H.	107	3.5 1.7 +2.7	128	4.2 1.4 -3.6	160	2.0 1.7 -1.0	182	1.2 1.0 +.1	240	4.2 2.2 +3.9
Ri.	95	2.3 1.9 +2.5	144	2.9 2.3 -1.6	182	2.1 1.9 +.2	240	4.5 1.5 -.5	320	2.5 2.1 +2.7

TABLE II (Continued)

Ab.	95	3.2 1.0 +3.1	128	5.5 1.4 +5.5	160	3.1 .9 +3.1	182	1.3 .8 +1.7	256	1.4 .9 +.9
Mc.	95	1.7 .5 -1.6	120	3.9 1.3 +3.8	160	1.9 .8 -1.7	213	1.0 .9 +.4	256	2.6 1.2 +2.0
Br.	95	2.9 2.2 +2.7	144	5.6 1.0 +5.2	182	6.4 1.1 +6.4	256	3.7 1.0 +3.6	384	2.7 2.8 +1.3
D.	95	1.7 .9 -1.4	128	3.3 3.1 -.3	160	5.9 2.4 -5.9	182	6.3 6.6 -3.4	256	11.2 5.4 -6.5
Bh.	120	1.6 1.1 -1.3	144	2.4 1.6 +.9	182	3.0 1.1 +3.0	240	5.1 1.8 +5.1	286	2.5 1.4 +1.5
P	103.0		135.0		172.9		230.6		299.4	
Average in v.d.	E	2.3		2.6		2.4		2.8		3.5
	m.v.	1.5		1.3		1.3		1.8		2.0
	C.E.	2.0		2.0		1.7		2.1		2.5
	G.C.E.	+.8		+1.0		+.5		+.5		+.8
Ave. in per cent. of a tone	E	.18		.15		.11		.09		.10
	m.v.	.11		.08		.06		.06		.05
	C.E.	.15		.12		.08		.07		.06
	G.C.E.	+.06		+.06		+.02		+.02		+.02

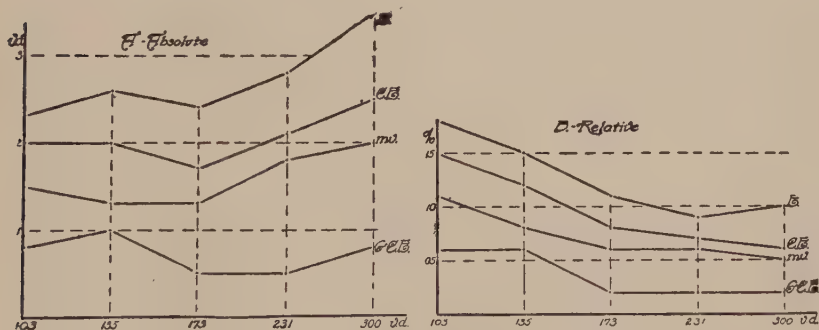


Fig. 1. Accuracy and the voice. (Table II)

is to be accounted for mainly by the fact that some observers were erratic on this tone, probably because the tone was higher than the observer commonly sings. As a matter of fact only half of the observers, nearly all of whom would be classed as bass or baritone in their range, show any tendency to decrease in accuracy at this

point; five show the tendency to increase in accuracy and the remaining four tend to remain constant.

This result is in harmony with results found by *Preyer* (16), *Luft* (13), *Meyer* (15) and *Vance* (26) on the sensory side, that pitch discrimination is approximately constant in terms of vibration frequency within this range. It is in harmony with the finding of *Berlage* (2) as quoted above: item 5 (second part), that average error diminishes with rising pitch if expressed in per cent. of vibration frequency; and item 6, with reference to the reproduction of one's own tones.

It is interesting to compare and to contrast these records with those of *Seashore* and *Jenner* (19), item 9, showing that the average error in the singing of a natural interval (third, fifth, and octave) varies approximately with the magnitude of the interval; (see also *Sokolowsky's* results above) and, item 10, showing that the minimal change is a relatively constant fraction of a tone within the octave.

The tendency for the C. E. to be in one direction (+) will be considered in a later section in connection with the constant errors in our other series of experiments.

EXPERIMENTS SERIES II: INTENSITY OF STANDARD

In the experiments of Series I, as stated above, two successive trials were made on each fork. Occasionally upon the presentation of the tones for the second trial at reproduction the observer would say "Let me hear that again; it sounds higher (or lower) than before", or "Is that the same fork?" Such remarks by careful observers led to this consideration of the intensity factor.

The same forks were used with the respective observers as in Series I. The tones were presented to the ear by the experimenter as before. But with half of the trials the standards were made about as strong as possible by striking the forks a heavy blow and presenting them near the ear. The other standards were made as weak as could be heard with distinctness. The observers were encouraged to sing with a medium volume of voice and not to imitate that of the forks, as is the natural tendency. Twenty records were made with each fork, ten on the "weak" and ten on the "strong" in the double fatigue order, as regards pitch and intensity. No successive trials were made on the same fork except on the highest and lowest. Having sung the tones from the highest the

observer would sing them in reverse order from highest to lowest; but a short pause was introduced between such successive reproductions. Of the eight observers tested, P, Ma., V. Z., C, An., V. H., S and Mi., the first six had no definite knowledge of the object in view.

The results are shown in Table III and graphically represented in Fig. 2. "*W*" denotes weak and "*S*" strong, while the other notation is the same as that previously used. It will be seen that the intensity of the standard tone has a decided effect upon the accuracy of reproduction.

(1). Increase in intensity causes a lowering in the pitch of the reproduction. The G. C. E. for S on each of the five levels is less than that measure for *W*., the minimum amount of difference being 1.4 v.d., the maximum 4.1 v.d. and the average for the five pitches, 2.3 v.d. In all the forty individual constant errors with the exception of two (see V. H.'s lowest tone and C.'s highest; in this latter "*W*." and "*S*." are just the same) the reproductions of the "strong" standards are lower than those of the "weak". If we compare these averages (C. E.'s and G. C. E.'s) with those of the previous series of experiments we find not only that the "strong" C. E.'s and G. C. E.'s are lower in the majority of cases than those of Series I, but that these measures for the reproductions made from the "weak" standards are somewhat higher than those of the former series. The effect of intensity in other words, is evident in both "weak" and "strong" standards, the former heightening the seeming natural tendency to sharp and the latter overcoming this tendency with a more powerful one to flat.⁷

(2). Strong standard tones cause general inaccuracy of voice control. Most of the observers stated that they were less sure with the "strong" standards. Others complained that the test made their ears tired. Reference to the mean variations and also to the E.'s and C. E.'s will show that in the majority of cases these amounts are

⁷ When the conditions of this experiment (Series II) were explained to P., the professional musician, he remarked off hand as he began the test: "Loud tones would make your nerves more tense and would in general tend to make you sharp." He was asked then and at other times during the test to let any conscious tendency to flat or sharp take care of itself *e.g.* not knowingly to correct for it. At the last P. said: "I am equally satisfied with my reproductions of weak and strong." Cf. P.'s record in Table III.

TABLE III. *Intensity of standard tones*

Ave. P.	105 v.d.		135 v.d.		174 v.d.		237 v.d.		301 v.d.	
	W.	S.	W.	S.	W.	S.	W.	S.	W.	S.
	E.	E.	E.	E.	E.	E.	E.	E.	E.	E.
	m.v.	m.v.	m.v.	m.v.	m.v.	m.v.	m.v.	m.v.	m.v.	m.v.
O.	C.E.	C.E.	C.E.	C.E.	C.E.	C.E.	C.E.	C.E.	C.E.	C.E.
P.	1.3 1.0 - .7	2.1 .6 -2.1	.7 .5 + .3	1.2 .9 -1.0	2.3 1.2 +2.3	1.4 1.4 + .3	1.1 .7 +1.1	.7 - .7 + .2	3.0 1.3 +2.9	1.0 .8 + .5
Ma.	1.1 .9 +1.0	1.1 .9 - .7	1.0 1.0 - .2	1.4 .8 -1.4	3.2 .8 +3.2	1.7 1.7 + .8	3.1 .6 +3.1	2.1 1.0 +2.0	2.1 .7 +2.1	1.4 1.0 -1.4
V.Z.	3.8 .7 +3.8	1.6 .9 +1.2	2.9 1.3 +3.1	1.1 .7 +1.0	3.2 .5 +3.2	.7 .7 - .2	1.9 1.5 -1.2	4.8 1.2 -4.8	2.0 1.7 + .3	1.8 1.3 -2.1
C.	1.2 .5 +1.2	.7 .7 + .2	1.9 .5 +1.9	1.1 1.1 + .3	1.7 .8 +1.7	.7 .7 - .1	3.0 1.1 +3.0	2.4 .9 +2.4	2.8 2.0 +2.6	2.6 1.0 +2.6
An.	6.4 6.4 +6.4	5.4 2.8 +4.8	1.9 1.3 +1.9	1.1 1.1 - .1	1.9 1.9 - .5	2.3 1.3 -1.9	3.2 1.2 -3.1	7.8 1.7 -7.8	3.7 1.8 -3.4	8.1 1.2 -8.1
V.H.	3.1 1.6 +2.6	2.9 1.7 +2.9	3.9 1.0 +3.9	3.5 1.5 +2.9	1.0 .8 - .7	1.9 1.8 -1.2	2.2 1.9 + .8	2.1 2.1 + .3	2.9 3.0 - .5	8.0 3.4 -8.0
S.	1.8 .5 -1.8	4.2 .7 -4.2	1.0 .5 +1.0	.7 .6 - .4	.8 .8 - .2	3.6 1.0 -3.6	2.1 1.2 -1.8	4.7 .9 -4.4	5.3 1.3 +5.3	5.5 1.4 +5.5
Mi.	9.3 1.9 +9.3	8.5 4.1 +8.5	1.5 .9 +1.5	1.7 .9 + .5	2.2 2.0 - .8	5.4 2.9 -5.4	4.0 3.1 +2.4	4.5 4.1 -2.1	7.8 1.9 +7.8	4.7 2.9 -4.6
Av. E.	3.5	3.3	1.9	1.5	2.0	2.2	2.6	3.6	3.7	4.1
Av. m.v.	1.7	1.6	.9	1.0	1.1	1.4	1.4	1.6	1.7	1.6
Av. C.E.	3.4	3.1	1.7	1.0	1.6	1.7	2.1	3.0	3.1	4.1
G.C.E.	+2.7	+1.3	+1.7	+ .3	+1.0	-1.4	- .5	-1.8	+2.1	-2.0

larger with strong standards, thus indicating conditions that operate against the best vocal control.

The matter of intensity has been considered in the field of pitch discrimination, where it must really be worked out. Seashore (21) makes the following statements concerning it:

"Extensive experiments show (1) that both trained and untrained observers may be influenced by intensity in their pitch judgment; (2) that although there is a tendency among the untrained, especially the ignorant, to judge the loud tone the higher, it may work either way; (3) that the same individual may show one tendency at one time and the reverse at another; (4) that for trained observers the two tendencies are about equal; and (5) that the tendency is more serious for large than for small intensity differences. Introspection shows that this confusion rests largely on motor tendencies, or motor images. We associate high and strong with strain—the reversal can in some cases be traced to a correction, conscious or unconscious, based on knowledge of this danger.

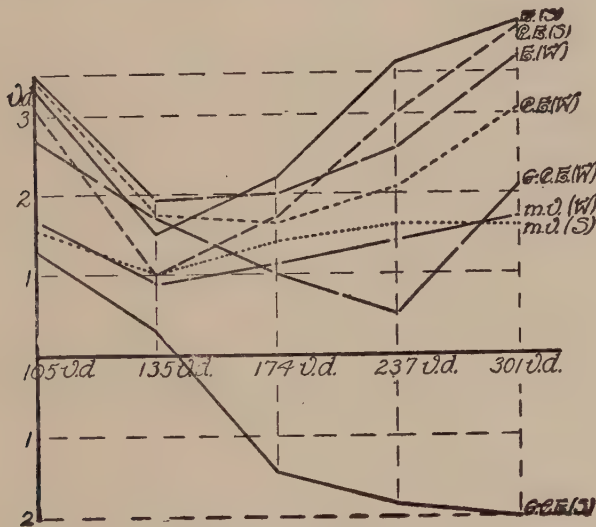


Fig. 2. The influence of intensity of standard tones.

"Experiments show that the just perfectly clearly perceptible tone is most favorable for accurate results. It is ordinarily purer than a stronger tone and favors concentration. Experimenters must guard against a very common tendency, usually unconscious, to facilitate the discrimination by making the tones loud; and untrained observers usually desire (unwisely) a loud tone."

These conclusions are found on tests made by Anderson (1) at the level of 435 v.d. Our results just stated led to a re-examination of the effect of intensity on pitch. Hancock (8) found that as

measured in terms of hearing alone the tendency to hear a relatively low strong tone as low is greater than is shown in this series for singing. All these facts make clear that in singing from a standard tone greater care must be exercised to keep the tone constant and at a most favorable strength. We have no adequate quantitative data to show what strength is best but the facts available tend to support the statement made by Seashore (21) that the just perfectly clearly perceptible tone is most favorable for accurate results.⁸

EXPERIMENTS SERIES III: VOLUME OF THE VOICE

The effect produced by varying the intensity of the standard tones suggested a parallel question concerning the relationship of voice volume and accuracy of reproduction. This problem was attacked in the following manner. The forks selected were the same for each observer as in the voice range test, Series I; they were presented to the observer's ear by the experimenter who endeavored to keep the intensity as nearly constant as possible, and the observer was instructed to reproduce the tones in three degrees of voice volume, "loud", "medium" and "weak". Ten trials were made on each fork with each of these three degrees of loudness of voice, the order being as follows: one trial on each fork from lowest to highest and after a pause from highest to lowest with "medium" intensity; from highest to lowest and back to highest with "loud" intensity; from lowest to highest and back to lowest on "weak"; highest to lowest and back on "medium" and so forth until the 150 trials were made.

These records are summarized in Table IV and represented in part in Fig. 3. In this table *L*, *M* and *W* represent respectively loud, medium, and weak, other notation is the same as in the foregoing tables.

Here we find again, as in the foregoing series, the tendency for accuracy in singing to remain a constant in terms of vibrations, except for the extreme notes, at which there is a decrease in efficiency, especially at the high note. The form of the average error curve (*E*)

⁸ The force of the blow changes the pitch of a fork, (See Winkelmann's *Akustik*, Vol. 2, p. 358) lowering it slightly, but this change in these forks could hardly be detected and certainly fails to account for the error in reproduction. See also Seashore (21).

here is entirely analogous to the form of the curve of pitch discrimination referred to above (16, 13, 15 and 26) but it represents a shorter range, as the voice has a shorter range than the ear.

The constant error for men here, as in the foregoing series, is in the direction of sharpening. It is a relatively constant fraction of a vibration for all pitches except the highest. The records for the medium and the weak tones practically coincide,⁹ and compare very favorably with those of Series I., but there is a uniform tendency to sing the loud reproduction highest. The average difference between the loud and the weak (see the G. C. E.'s) is here .6 v.d. This is not a contradiction, but the reciprocal of the results found in Series II: namely, that the loud (or strong) standard is reproduced low.

It will be remembered that in Series II the standard was made strong, the observer tried to produce a tone that subjectively seemed the same in pitch, and that practically all of his reproductions were flat. This result in the light of Series I, where sharpening was the rule, seemed to warrant the conclusion that the strong tone is judged low. Now in Series III we have a confirmation of this; here the standard is of medium intensity while the reproductions are varied: loud, medium and weak. It seems therefore that the instant the observer commences his loud reproduction he is subject to the same error in judgment as was revealed in Series II, and that to make his reproduction subjectively equal in pitch to the standard, he thinks it necessary to raise. This brings about abnormal sharpening: the average G. C. E. of Series III is +1.3 v.d. as against +.7 v.d. for Series I, where intensity differences were at a minimum.

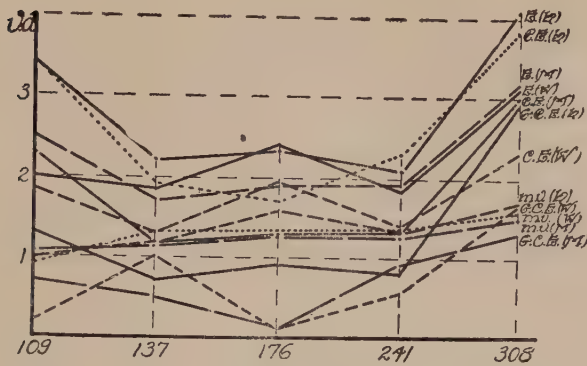
The agreement of the errors (G. C. E.'s) in these two series (II and III) at once offers an explanation for them: the error is primarily one of hearing which is basal and the chief cause for the error in singing.¹⁰ This is in harmony with the contention of Klünder

⁹ Medium and weak tended to be confused by the observers who would frequently have to be reminded that they were not making sufficient difference between them. This would imply that they each seemed more natural and less distinct than the loud, which is borne out by the fact that the average G. C. E. for weak (Series III) is identical with that for Series I, i.e. +.7 v.d. However it should be noted that the curves for weak are less regular than in Series I.

¹⁰ It is interesting to note that in Series II the high strong is flatted most, while in Series III the high loud is sharpened most.

TABLE IV. *Volume of the voice and accuracy*

Ave. P.	109 v.d.			137 v.d.			176 v.d.			241 v.d.			308 v.d.		
	L E.	M	W	L E.	M	W	L E.	M	W	L E.	M	W	L E.	M	W
	m.v. C.E.			m.v. C.E.			m.v. C.E.			m.v. C.E.			m.v. C.E.		
O.	3.0	2.0	1.3	2.8	1.4	.9	.8	.9	1.8	1.9	1.3	1.5	1.2	2.7	4.1
P.	.6	.8	.7	.5	1.2	.5	.8	.9	.9	1.1	1.3	.6	1.1	1.6	.8
	-3.1	-1.9	-1.2	-2.7	-.9	+.1	-.3	-.2	+1.6	-1.7	-.5	+1.5	+.8	+2.4	+1.4
Ma.	2.0	1.8	1.3	2.0	1.4	1.7	2.5	3.0	3.9	2.2	1.0	1.1	1.4	1.7	2.4
	.7	1.0	.5	1.6	1.4	1.5	1.3	1.3	.9	1.0	1.0	1.1	1.2	1.5	1.3
	-2.0	1.5	-1.3	-1.5	+.1	-.5	+2.4	+3.0	+3.9	-2.2	+.5	+1.1	-1.2	-1.2	+2.3
VZ.	3.9	3.5	3.2	2.9	2.5	2.6	2.3	2.0	1.3	2.2	2.5	2.1	1.0	1.2	1.3
	1.0	1.1	1.4	1.1	1.2	1.4	1.6	1.9	.7	1.2	1.0	.7	1.0	1.2	1.0
	+3.9	+3.5	+3.2	+2.8	+2.5	+2.3	+2.0	+1.8	+1.0	-1.7	+2.3	-2.0	-.5	-.4	+1.2
S.	2.3	2.3	2.7	1.2	.9	1.1	3.2	2.4	2.6	.9	1.4	.9	7.6	5.2	5.2
	.8	.4	.5	1.2	.8	1.0	1.1	.6	.7	.8	.9	.8	1.9	1.4	1.5
	-2.2	-2.2	-2.6	+.1	-.5	+.6	+3.1	-2.4	-2.6	+4.4	-1.0	-.3	+7.5	+5.2	+5.2
C.	1.4	1.2	1.0	1.9	1.4	1.7	1.5	.4	.7	3.0	3.6	2.4	5.6	3.9	2.8
	.5	.6	1.0	1.1	.6	.8	1.5	.4	.8	.9	.4	1.1	1.5	1.6	1.1
	+1.4	+1.2	-.6	+1.3	+1.4	+1.7	-.5	-.4	+.4	+3.0	+3.5	+2.3	+5.5	+3.8	+2.8
VH.	9.0	4.9	2.5	2.1	1.7	1.6	2.9	2.2	4.3	1.6	2.2	2.0	2.2	4.2	3.0
	1.8	2.1	1.3	1.4	1.5	1.6	2.3	2.0	3.6	1.7	2.8	2.0	1.5	2.4	2.8
	+8.9	+4.9	+2.5	+2.1	-.8	+2.0	-2.1	-1.2	-2.5	+.4	+.8	-.4	-1.3	-4.2	-2.0
Mi.	2.3	1.6	1.7	2.7	2.6	2.7	3.1	2.3	2.0	3.1	1.3	2.6	9.5	3.5	3.0
	1.6	1.5	1.2	1.5	1.6	2.0	2.4	2.3	1.6	2.3	1.4	2.2	3.5	1.0	2.6
	+2.2	+1.2	+1.3	+2.7	+1.9	+2.3	+1.5	+.3	-1.2	+3.1	+.7	+1.8	+9.5	+3.5	+1.2
Av. E.	3.4	2.5	2.0	2.2	1.7	1.8	2.3	1.9	2.4	2.1	1.9	1.8	4.1	3.2	3.1
Av. m.v.	1.0	1.1	.9	1.2	1.2	1.3	1.6	1.3	1.3	1.3	1.3	1.2	1.7	1.5	1.6
Av. C.E.	3.4	2.3	1.8	1.9	1.2	1.3	1.7	1.3	1.9	2.3	1.3	1.3	3.8	3.0	2.3
C.G.E.	+1.3	+.7	+.2	+.7	+.5	+1.0	+.9	+.1	+.1	+.8	+.9	+.6	+2.9	+1.3	+1.7



(12) that the ear is the chief criterion for regulating the voice. But the result quoted from Hancock (8): that the hearing error is greater than the singing error (when dealing with a low, strong tone), together with the fact (Series III) that there is relatively more flattening with a strong standard than there is sharpening with a loud reproduction, would conform to the conclusion reached by Stern (24) that the kinaesthetic sense of the singer is also an important factor.

One would expect a larger mean variation (m.v.) for the tone that has the largest error, but the table shows the mean variation to be practically equal for all three intensities of sound. This may be taken as a mark of the relative constancy of the motive for the intensity error.

The agreement and the remarkable uniformity in these two laws as shown in Series II and III would indicate that we are here dealing with an important factor of which we must take cognizance, both in the hearing and the producing of musical tones.

EXPERIMENTS SERIES IV: TIMBRE OF STANDARD TONES

Kländler (11 and 12), Cameron (4) and Sokolowsky (22) in their researches used organ tones for standards, while Berlage (2) made use of tones from the voice. Having ourselves used tuning forks it seemed advisable to ascertain if timbre differences in the standards affect the accuracy of reproduction.

The standards selected for the test were: a large disc tuning fork (144 v.d.) sounded before a resonator, the dichord (137.5 v.d.) energized by bowing, and an organ pipe blown by mouth. In using the latter, because of the variability of the blow and hence the uncertainty of the pitch sounded, the vibration frequency of each standard tone was read on the tonoscope and entered in a parallel column with the reproductions. The tones were so far as possible of uniform intensity, they were sounded for approximately 2 seconds and after the interval of 1 second reproduced on *a*, as in "law" with medium volume of voice. Twenty trials were made on each standard, and because the effect of timbre was the point of interest, the reproductions were in groups of five successive trials, the standard of course being sounded before each attempt.

TABLE V. *Timbre of standard tones and accuracy*

O.	Fork 144 v.d.			String 137.5 v.d.			Pipe Av. 150 v.d.		
	E.	m.v.	C.E.	E.	m.v.	C.E.	E.	m.v.	C.E.
P.	2.5	.6	+2.5	1.8	.5	-1.9	.9	.4	-.1
S.	1.0	.9	-.8	1.9	.8	-1.9	1.1	.5	-.3
Ma.	1.2	1.3	+1.2	1.2	.5	+1.2	1.3	.8	+.6
V.Z.	1.3	.5	+1.2	2.1	.4	-1.9	.6	.5	-.5
Mi.	2.0	.6	+2.0	.6	.5	+.3	.5	.4	-.2
Av. E.	1.6			1.5			.9		
Av. m.v.		.8			.5			.5	
Av. C.E.			1.5			1.4			.3
G.C.E.			+1.2			-.8			-.1

The results of this series of experiments are summarized in Table V. Judging by the magnitude of the average error and the constant error, the record is in favor of the organ pipe. This is probably due to the fact that this tone is most nearly like that of the human voice in tone-color, or timbre. The introspections of our observers, all of whom have good musical ability and were practiced in observation, are however not in accord with this. Four of the five stated that the string was the easiest standard to imitate. P, the one professional musician in the group, felt that he did best on the fork. But reference to the table shows that it was here that he made his largest errors and even the largest made by any observer on that standard. S. stated that the string was by far the best as a standard but made his smallest errors, and the smallest made by anyone, on the fork. It must be noted also that S. has had more practice with forks than any other member of the group. Practice

is undoubtedly a factor and the value of it for a particular observer depends chiefly on what associations are awakened by a given tone-color. Purity, for example, may be thought of as thinness, and secondarily as highness of tone. While tuning forks, being relatively pure and free from over tones, are at a disadvantage on the side of richness, it is also true that in most groups the observers are about equally unpracticed in singing with forks, which is an advantage from the standpoint of measurement. The forks also are decidedly more constant in pitch than any other type of standard tone. Two of the observers noticed a tendency to imitate the timbre of the standards.

From the above observations it seems fair to conclude that richness of tone favors accuracy in the reproduction of any particular standard.¹¹

EXPERIMENTS SERIES V: VOWEL QUALITY AND ACCURACY

Berlage (2) introduced the problem of the influence of vowel quality (or change in the timbre of the singing voice) upon accuracy in imitating pitch, and made measurements on this point for the purpose of determining the effect of mouth resonance upon the pitch of the reproductions. In considering the problem there is no thought of discrediting the results found by Berlage. The tonoscope method of recording has enabled us to take many more records than were used by him in computing his results and the matter is of such far reaching importance that it seemed worth while to include in our study a series on this factor, limiting our measurements to the following vowels:

- u* as oo in "toot"
- o* as o in "no"
- a* as in "ah"
- e* as e in "there"
- i* as i in "machine"

In addition to these, humming the tone was introduced in the test as the "hum" seemed to have no marked vowel quality.

¹¹ Starch (23 p. 52) in his conclusions on the effect of timbre in the localization of sound makes this statement: "The richer and more complex a sound the more accurately it can be localized."

Three forks of the large disc variety were used as standards, the pitches being: 144, 182 and 240 v.d. and each of these three tones was reproduced to the five vowels and the "hum" twenty times, a total of 360 trials for the individual observer. The test was divided between two equal periods. The order of reproducing was two trials on each fork to each vowel in the double fatigue order, illustrated as follows: 144 to *u*, 182 to *u*, 240 to *u*, pause, 240 to *u*, 182 to *u*, and 144 to *u*; then 240 to *o*, 182 to *o*, etc., followed by 144 to *a*, 182 to *a*, and so on throughout the test, the order of the vowels being *u*, *o*, *a*, *e*, *i*, and "hum." All standards were presented to the ear for a duration of approximately 2 seconds and an interval of 1 second was allowed before the singing.

That the vowel quality is a factor influencing the accuracy of reproduction is borne out by the results of the series as shown in Table VI. The average error (E) and the mean variation (m.v.) are given merely for an index to the reliability of the record. They are both large as compared with the constant error (C.E.) which is the factor in terms of which we desire to measure the effect of vowel quality on reproduction.

Although there are characteristic differences for the three pitch levels and for the different individual observers, the results in Table VI may be fairly represented by a single curve (Fig. 4). This

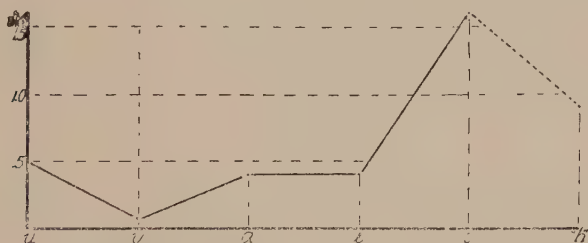


Fig. 4. Vowel quality of the voice and accuracy (Table VI).

shows graphically the algebraic average of the records (G. C. E.) for each of the vowels and for the three levels, 144 v.d., 182 v.d. and 240 v.d. There is a tendency for the vowels to fall into three groups: namely, (1) *o* sung the lowest, (2) *a*, *e* and possibly *u* sung moderately sharp and (3) *i* sung decidedly sharp. These facts would seem to point to the general conclusion that the higher the dominating overtone in a vowel clang, the higher that vowel will be sung. In Fig. 4, *u* offers the single exception to that rule.

The hum was supposed to be neutral as it was moderately weak and the record was made from the nasal breath. This assumption is confirmed by the record which gives the hum a middle place with *a* and *e*.

It must be remembered that there is, in the foregoing records which were sung on *a*, a tendency to sharp by about the amount of sharp for *a* here. That tendency is probably due to some other cause than timbre. It may therefore be suggested that *a* and *e*, the vowels usually sung when one is free, are fairly neutral; *o* (and possibly *u*) are sung relatively flat and *i* relatively sharp. This view, it will be observed, is confirmed by the hum.

Our results seem to differ radically from those of Berlage (4), second part of item 4, in the observations which are common to both. But our method also was radically different; moreover, his conclusion (item 4) is somewhat modified when we read in his article p. 76 where the results of the vowel experiments are discussed. "Accordingly one may look upon a slight increase in the variable error as probable with vowel change" (*i.e.* when the observer tries to reproduce his own pitch but on a different vowel) . . . "other generalities cannot be deduced for the table . . ."

These results in reference to vowel quality are of so far reaching significance for speech and song that we may not venture further discussion for the matter must be made a special object of investigation for verification of the empirical data and in search of an interpretation. It seems safe however to proceed in our work using "*a*" as the vowel quality for reproductions.

EXPERIMENTS SERIES VI: ACCURACY IN SINGING

Having gained some insight concerning the influence of voice range, standard tone intensity, voice volume, standard tone timbre, and voice timbre, on the accuracy of voice control, we now turn to the main problems of our research. These may be restated as follows: (1) What is the average error of the human voice in reproducing the pitch of a tone? (2) What is the average minimal producible change of the voice? (3) Is there any general tendency to sing sharp or flat? (4) How does the average performance of men and women compare on the above three points? All the studies referred to in the historical account contain results which cast light on some of these problems. But in almost every case these

results and problems are secondary to the main interest of the study; and moreover the number of observers and observations is usually quite limited. In Series VI therefore, we have made these problems the central issue on a large group of persons to give our results significance as norms.

Apparatus and method

Standards. With the aid of the tonoscope, eleven large disc forks were tuned to the following pitches: 128, 128.5, 129, 130, 131, 133, 136, 140, 145, 151, and 158 v.d. The series of pitch increments between the forks was therefore: .5, 1, 2, 3, 5, 8, 12, 17, 23, and 30 v.d. as measured from 128 v.d. This series of tones was used for men. For the women a second set was provided on 256 v.d. as a basis, namely, 256, 256.5, 257, 258, 259, 261, 264, 268, 273, 279, and 286 v.d. In this second set it will be noted that the same pitch increment (absolute) were used as in the 128 v.d. set instead of the relatively equal increments. In this respect the procedure was based upon the conclusions reached in Series I.

Koenig resonators were provided for each set of forks. As the increments were small it was found that one resonator would speak sufficiently well to several tones. In the case of the 128 set three resonators were used: first, 128 v.d. to and including 136; second, 140 and 145 v.d.; and third, 151 and 158 v.d. For the higher set two resonators were found sufficient: first, 256 v.d. to and including 268 v.d.; and second, 273, 279, and 286 v.d. Both series of forks as reinforced by the resonators gave tones of pleasing quality and medium intensity.

Observers. Two hundred and one individuals, ninety-four men and one hundred and seven women, took the test which is about to be described. This number comprised those enrolled in the elementary psychology courses in the University of Iowa, 1912-1913. Of these about one hundred fifteen were sophomores; the remaining were upperclassmen. None of them had had any practice in this test. Among them were some excellent vocalists and some others who claimed never even to hum or whistle and to have difficulty in recognizing old and familiar tunes if unaccompanied by words. No one was excused because of his inability and no one was selected because of ability, for it was desired in so far as possible to secure what might be considered an average group. A previous lecture on

the measurement of musical capacity had successfully aroused the interest of the observers so that they entered into the experiment with zest, many of them desiring to secure their individual results.

The charge. The instructions were given by word of mouth to each person, although the appointments were so arranged that one observer was present while another was taking the test and so became familiar with the procedure before he actually entered upon it. Supposing the observer to be a man the instructions would be as follows:

"Mr. ———, we have here a series of eleven tuning forks. This one (striking the 128 v.d. and presenting it before the resonator) is c below "middle c", it is a tone of 128 v.d., the lowest tone in the series; we will call it "o". This one (striking and presenting the increment fork 30, 158 v.d.) is considerably higher than o as you easily notice, and is the highest one in the group. These other forks all represent pitches between the two which we have sounded. The test to-day consists in singing these eleven tones one after the other as they are given. They will be presented in pairs. First we will sound the o, the lowest one of the tones; you will listen carefully to it and then sing a tone of the same pitch. Immediately after your singing, the highest tone in the group (30, 158 v.d.) will be sounded; you will listen and sing that one. Then the o will be sounded again and, after you sing it, there will come the next to the highest tone (23, 151 v.d.); and so on we will come down one step at a time always reproducing the o before each of the interval forks. When you have tried all the tones in the series you will go back over them in the reverse order. Simply imitate as nearly as possible the pitch of each tone as it is given, always remembering that the o is the lowest one in the series. Sing all the tones with a natural voice volume and use the vowel "a" (a as in "ah") and whenever you feel dissatisfied with any trial ask for a repetition."

Following these instructions, in order to put the observer at ease and to satisfy his curiosity, the experimenter gave a brief explanation and demonstration of reading on the tonoscope.

The test. The forks were presented to the resonator by a helper who gave his attention solely to the task of sounding the tones in the right order and with as nearly uniform intensity and duration as possible. The tones were sounded with medium intensity varying towards the "weak." The observer sat on a high stool or stood at

the side of the instrument in a position which kept him from seeing his own record. He sang the tones into a metal speaking-tube placing the lips lightly against the fingers of the hand which grasped the mouthpiece. The arm was supported by an adjustable rest and, so far as could be, strain and unnaturalness were avoided.

A few preliminary trials were given on increment 0-30 in order that the observer might find himself becoming somewhat familiar, not only with the tonal range covered by the standards but with the experience of taking pitch from a tuning fork. The series was then given in pairs in the following order: 0-30, 0-23, 0-17, 0-12, 0-8, 0-5, 0-2, 0-1, 0-.5; 0-.5, 0-1, 0-2, etc., back to 0-30. The complete test consisted in singing the series thus five times. This gave one hundred reproductions of the 0, and ten on each of the increment tones, a total of two hundred tones for each observer. This series therefore contains forty thousand records. The test as outlined could not be performed with care in less than 30 minutes. In some cases and especially with non-musical persons a much longer time was required than this.

Throughout the test we endeavored to keep the observer seriously trying to sing the exact pitch of the forks. To this end it was deemed desirable to offer some encouragement, especially during the first fifth of the experiment, no matter how poor or good the record. It was observed that encouragement did not cause the singers to be self-satisfied or careless but rather served to make them try the harder. It helped moreover to create an atmosphere of ease and naturalness. But while there was encouragement there was also some criticism. If, for example, the observer was singing the 0 flat 5 or 6 v.d. regularly he was told to listen more carefully to the standard and to make sure that he had the right pitch but no intimation was given as to the character of the error. Little rest periods of twenty seconds were rather frequent and were found to be of much service. Many times it was noted that after such a period the errors were decidedly smaller than before.

A few questions concerning the observer's musical education, voice range, and ability to play and sing were asked during or following the experiment and the answers together with some comments regarding his performance of the test were made matters of record.

Justification of procedure

Before considering the results of this series it remains to justify the form of procedure as outlined above in the light of the sources of error revealed by our previous experiments and by those of other investigators.

Voice level of test. Our experiments (Series I) on the accuracy of pitch singing within the voice range demonstrated that the errors are relatively smaller on the higher tones. Unpracticed observers however, will much more readily try a tone that is medium or low than one that is high. It therefore seemed best for general testing to choose a voice level which all would recognize as being well within range. The selection of 128 to 158 v.d. for men and of 256 to 286 v.d. for women is thus the result of considerable experience in testing groups of individuals, and seems further justifiable on the grounds of pitch discrimination as previously stated.

Forks for standards. Tuning forks were retained for standards even though the records of Series IV indicate that the organ pipe and dichord can be imitated more accurately. Forks are very simple, easily manipulated, of practically constant timbre, and at the same time reliable in pitch. And if, as in our test, a series of tones differing from each other by slight degrees of pitch is desired to be sounded in rapid succession, tuning forks are the most reliable apparatus. Furthermore they are used so little for general musical purposes that in testing with them no group of observers is given undue advantage.

Many standards *vs.* one standard. Berlage (2) found that his observers could reproduce their own voice tones more accurately than tones given by some one else, the increase of precision showing itself chiefly in a decrease of the constant error. We have frequently noticed a tendency, which is a corollary of Berlage's conclusion. Observers when making successive trials on the same standard very often reproduce their own reproductions rather than make new efforts at imitating the real standard. The observer finds it much easier to reproduce his own previous tone, duplicating the muscle tension and mouth resonance which he experienced at that time and felt to be satisfactory. Indeed, even though he conscientiously work against this tendency, he can not overcome it entirely if engaged in making successive trials where the pauses between are brief. This is confirmed by the fact that frequently when observers

for some cause or other have been dissatisfied with attempts and desired new trials giving them immediately they would in the new trials unconsciously repeat the identical pitch given before. This same tendency has sometimes been evident even in large and unusual errors which the experimenter might rule out, asking for new trials. In view of these considerations it seemed best in our general test to adopt the principle of many standard tones and no successive trials on the same tone.

The increments between the forks were made small and of varying magnitudes for two reasons: first, in using these small increments we do not complicate our work with the factor of musical intervals, and second, in using a series of small increments we make possible the measurement also of the ability to make faint shadings (sharp or flat) in the pitch of the voice. The selection of increments is arbitrary. These particular steps were chosen because they have been found satisfactory in work with pitch discrimination (21) at the level of 435 v.d. and, as before stated, extensive research by Vance (26) and others shows that pitch discrimination is practically constant in terms of vibration frequency in the middle range of tonal hearing here covered. This is also the ground for making the increments for the women the same number of vibrations instead of the relative parts of a tone, in which case they would have been doubled.

Sounding the two tones. Seashore and Jenner (19) employed the method of "least producible, or minimal, change". The observer sang the standard or a tone at a given interval from it and then reproduced his own reproduction, save that he made it "the least possible" sharp or flat according as the experimenter might direct. While this will undoubtedly become a standard method in extensive work with an observer it is not suited to tests of a single sitting, first, because ability is rapidly improved by practice and, second, because the observer tends to be easily satisfied with his effort. The better way is not to rely on the changing subjective standard of the observer but to provide a series of constant objective increments and give him the opportunity to find his own level as by the method of constant stimuli in lifted weights or pitch discrimination. Such a series has been provided in the standards and increments mentioned above.

Order of standards. Manifestly the standards might be presented

to the observer in any one of a number of different orders. After trying out the matter thoroughly with the help of three good observers we selected the order of presentation above described for the following reasons: (1) to give the tones in pairs (0-30, 0-23, etc.) takes direct advantage of all the latitude which the series provides. Most observers can easily detect the difference 0-30, while many (theoretically about 25 per cent.) would be baffled to find a difference between 23 and 30; (2) to begin with the largest increment and work towards the smallest has the double advantage of establishing confidence in the attitude of the subject and of stimulating effort; (3) to give the increments in a series and in double fatigue order rests the voice from the unusual strain of making the least producible change, and (4) to explain definitely at the beginning of the test that all the increments are in one direction, *i.e.* above the 0, simplifies the problem and puts it more definitely under control than if uncertainty as to change of direction in standards were allowed. The test is therefore not to measure the judgment for direction of pitch difference but the judgment and expression of the amount of pitch difference between two tones. In pitch discrimination it is well known that much depends upon the direction of the expectant attention.¹² And should we present the standards of our test in a chance order we would complicate it exceedingly at the critical point of the smallest increments.

Time intervals. At the very beginning of the test the intention was to allow an interval of 1 second between the breaking off of the standard tone and the singing by the observer. But the method

¹² An idea of the influence of this same source of error operating in the field of singing may be gained from the following illustration. The author in instructing a very fine observer thoughtlessly said (the error was altogether unintentional); "We have here two forks, the first, 128 v.d., and the other one 3 v.d. higher, 131 v.d. You will please sing them one after the other. I will give the lower one first." Then the forks were presented and reproduced as directed. When we came to the twelfth trial the observer remarked: "I seem to feel strain to bring the 131 v.d. up." In the moment of reflection following this remark the writer recognized that he had made a mistake in instructing the subject, as the so-called "131 v.d." was really 3 v.d. lower than 128 v.d., or 125 v.d. We find in the twelve trials made that the average reproduction of 128 v.d. is 123.6 v.d. while the average pitch given for the supposed 131 v.d. (really 125 v.d. is 124 v.d. The misunderstanding and therefore expectant attention changed the direction of the reproductions, and brought in much larger constant errors than are usual for this individual. It should also be noted that the errors are minus.

was soon given up as, in this case cumbersome and unpractical, and furthermore we did not care to complicate our test with the factor of tone memory. (See *Berlage* (2) and *Sokolowsky* (22)). The observers in their usual singing with musical instruments make no such perceptible time intervals. They sing with the tones of the instrument, perhaps holding them somewhat longer than is done by the instrument. When the standard has been sounded the attention is centered, the muscles of the larynx almost involuntarily assume a particular tension and it is unnatural to wait for the beating of a metronome or some other signal to begin singing. If the unpracticed observer is told to make his own interval, unless checked up diligently, he will very soon be making intervals that are exceedingly short, if indeed he is not singing simultaneously with the standard tones. The method followed therefore was to sound the forks for approximately 1 second, encouraging the observer to begin his tone during the sounding of the fork and to hold it longer than the fork.

It may be objected that one might sing fairly accurately judging simply on the secondary criterion of beats between his voice and the standard tone. *Helmholtz* indeed (9 p. 326) suggests this as a convenient method for the singer to use for checking his own accuracy in practice exercises. While it would be possible for a highly practiced observer it can hardly have much influence in our test. The author made it a point to question frequently regarding the way observers judged of their success in reproducing tones and was not able to find any one who knowingly made use of this criterion. It is however quite possible that the roughness of 6 or 8 or more beats per second may occasionally have caused some observer to be dissatisfied with his attempt. But the tonal fluctuations and adjustments which are necessary to bring about a lessening of the frequency of beats between the voice and an outside standard are easily recognized with the tonoscope; no such "finding" process was observed.

Another time interval which must be considered is that between the *o* (128 or 256 v.d.) and the increment fork of any particular pair. In order that the standards for minimal change of voice may have their greatest value the interval indicated must be as short as possible admitting of a quick, direct comparison of tones; otherwise the test practically resolves itself into the singing of a single

tone. Hence the presentation of the increment fork followed immediately upon the close of the observer's reproduction of o, the subject being encouraged to make the reproduction about 1 second in length. The increment forks were struck while the o was being reproduced. This was, however, no distraction as only a slight blow on the practically noiseless sounder was necessary, and the forks could not be heard until presented before the resonators. Following the reproduction of each increment fork there was a period of about 2.5 seconds before the next sounding of o.

Other factors. In the matter of intensity of standard and intensity and vowel quality of the voice we took direct advantage of our previous work and adopted such conditions as would give the most normal results according to those findings. By the use of resonators at a considerable distance from the observer's ear we found a satisfactory means of controlling the intensity of the standards,¹³ while the intensity of the voice had to be judged subjectively and watched by the experimenter. And in the selection of "ä" we are using that vowel quality which according to Berlage and our own results affects least the constant error of the reproductions.

Tables of data

The constant error (C. E.) and mean variation (m.v. were found for the ten trials on each fork of the ten pairs given in the test. These twenty C. E.'s. and twenty m.v.'s for each individual tested are embodied in Table VII, which has been divided into two parts, A. and B., for the men and women respectively. In the first column of the table, at the left, are given the numbers which stand for the individual observers. This numbering is in no sense a ranking, but simply for convenience in handling the data and aid in identification. Odd numbers are used throughout to refer to women and even numbers to men. The second column from the left shows the C. E. and m. v. (the latter is under the former) for the ten trials on o. when used in the pair o-30. The same measures for the ten trials on variant (or interval) tone 30 are given in column three, and each of the successive smaller increments are represented in the same manner. The arithmetic averages for the constant error

¹³ It is possible that the pitch of a standard is not only varied by its intensity but also by its position when held near the ear.

(C. E.) and the mean variations (m.v.) on both standard and variant are presented in the two columns headed O. V. (Arithmetic). The algebraic averages for the constant errors (C. E.) on both standard and variant are given in the two O. V. columns at the extreme right of the table.¹⁴

The consolidated footings for Table VII, section A and B, are given in Table VIII. The top notation is thus the same as in Table VII. A contains the final footings for men and B for women. The footings are set out as follows: Ave. m.v. is the average mean variation for the respective points in terms of vibrations; C. E. % +, the per cent. of individuals who made a constant error in the direction of a sharp; % —, the per cent. of those who flatted; % O, the per cent. of those who made no appreciable constant error in the ten trials; C. E., v.d., the average magnitude in vibrations of the constant errors, without regard to sign; and G. C. E. v.d. the tendency of the constant errors for the group, the algebraic mean. At the right, the grand averages for both groups are presented under the headings designated above.

Comparison of the abilities of men and women

The most striking general feature of these experiments is the fact that women show the same ability as men, vibration for vibration, although the women sang an octave higher than the men. The data on which this assertion is based may be traced most readily in the curves, Figs. 5-8, 10. In Fig. 5 C. it is seen that the curves for the average constant errors on the standard as well as on the variant practically coincide. On the standard they are almost straight lines, the variation for the men being from 1.36 v.d. to 1.66 v.d. with an average of 1.54 v.d., while in the case of the women the variation of this measure is from 1.52 v.d. to 1.81 v.d. with an average of 1.65 v.d. The curves for the variants do not come so near coinciding; they are of the same form, but the women have the advantage, their range of C. E. falling between 1.69 v.d. and 6.71 v.d. with an average of 4.86 v.d., while that for the men lies between 2.59 v.d. and 7.15 v.d. with an average of 5.32 v.d. As further confirmation of the fact that the average con-

¹⁴ There would be little gained by placing E., the crude error, in our table as this measure is something of a cross between C. E. and m.v. and serves simply to indicate the distribution of the constant errors.

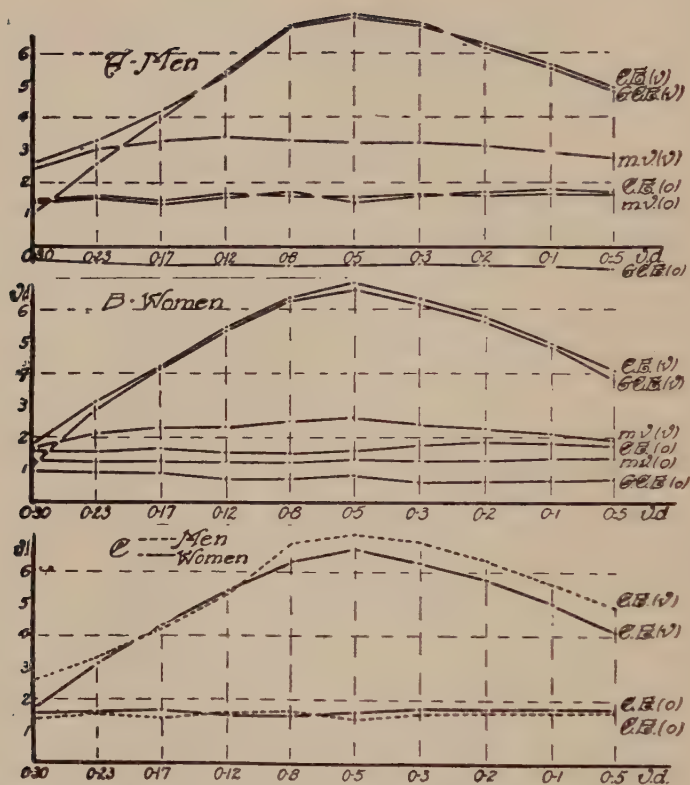


Fig. 5. The data in Table VIII. If there had been no errors all curves would coincide with the base line. The amount of deviation is indicated at the left in terms of vibrations; the increments on the base line. O denotes the standards (128 v.d. for men and 256 v.d. for women); V the variants; C.E. average (arithmetic) constant error; G.C.E. the algebraic constant error or general tendency of the group; and m.v. the mean variation. G. C. E. above the base indicates plus or sharp and below minus or flat.

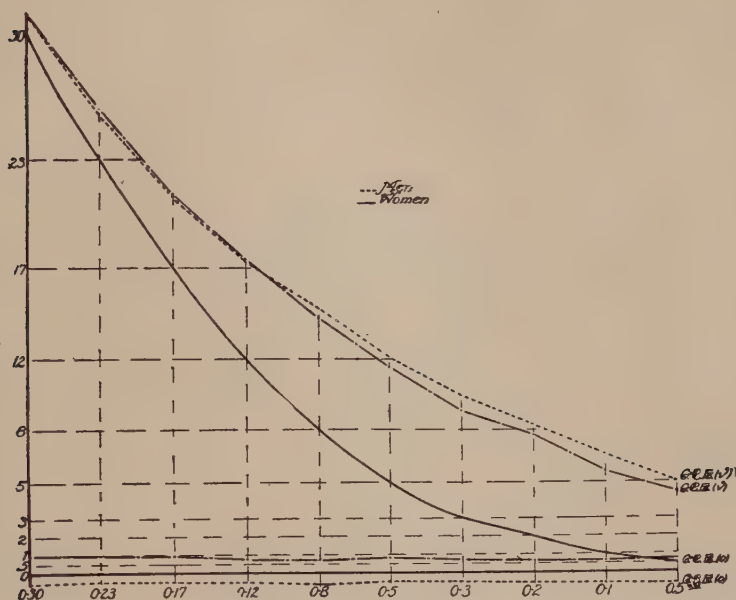


Fig. 6. Intervals as sung. (Table VIII). The distribution of the group constant errors (G.C.E.) for the standards (128 and 256 v.d.) and the variant in each interval. The intervals represented by the forks are shown in the heavy solid curves with which the other curves would coincide were there no errors in singing.

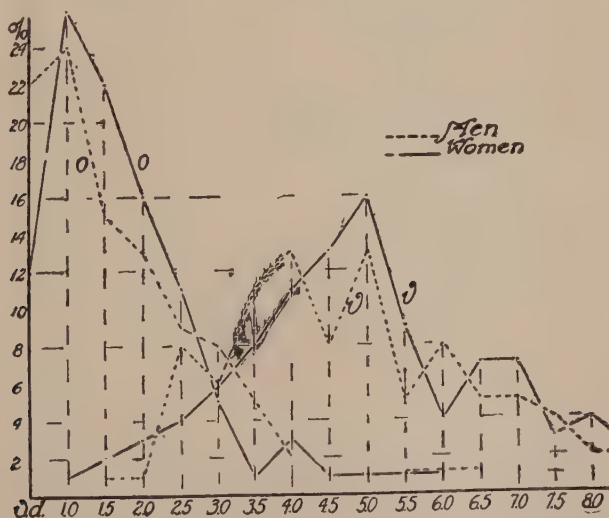


Fig. 7. The distribution of the average constant errors of all intervals for each observer with reference to the magnitude of the error. The data for this figure are found in the columns headed Arithmetic Average in Table VII. O, the standard tone: V, the variant.

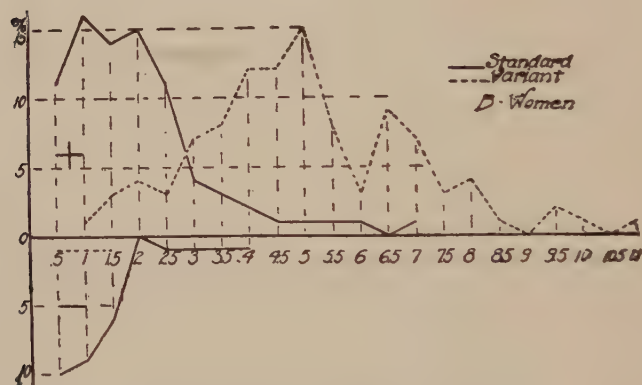
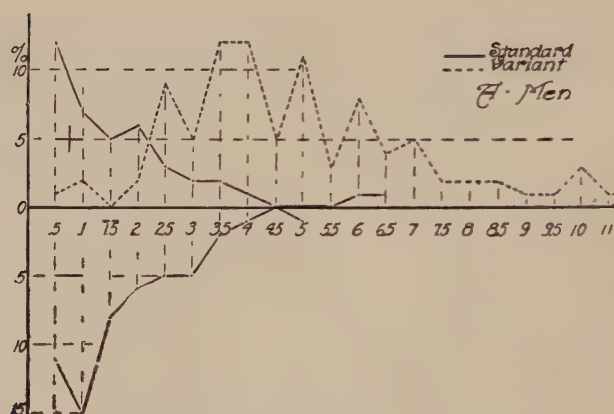


Fig. 8. Distribution of constant error, flat being denoted by — below the base and sharp by + above.

stant errors for both men and women represent approximately equal magnitudes attention is called to Fig. 7 in which is presented the distribution of the average constant errors of all intervals for each observer with reference to the magnitude of the error. The men have a slightly better record on the O. but the women have a more than compensating advantage on the V.

A corresponding agreement in the records for men and women is seen also in the constant tendency for the group (G. C. E. Fig. 5 A and B, 6, and 8 A and B). While the women tend to sharp and the men to flat on the standard (see Fig. 6) the amount is not far from equal in the two cases. (Cf. Table VIII, 65 per cent. of men flat on O while 67 per cent. of women sharp). In view of the general tendency of both men and women to sharp on the variant this difference in the tendency on the standard gives an advantage to the women as regards accuracy in the singing of the interval. An advantage which amounts to an average of over 2.0 v.d.

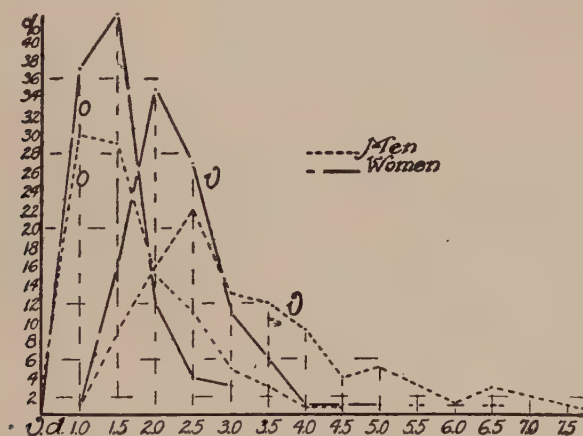


Fig. 9. The distribution of the mean variation (m.v.) for individuals (Table VII, average m.v. at right) with reference to the magnitude of the variations. O, m.v. of the standard tone; V, m.v. of the variant.

In the mean variation (Figs. 5 and 9), which is an important criterion, the advantage is more clearly in favor of the women, particularly in the singing of the variant. There are more men than women with a relatively large variation: but the mode in the case of O is slightly better for the men than for the women. The averages

of the men (Table VIII) are 1.54 and 3.05 v.d. as against 1.29 and 2.21 v.d. for women.

Taking all the data into account the general balance of all scores results practically in a draw:¹⁵ men and women sing with equal accuracy (in terms of number of vibrations of error) although the former sing at 128 v.d. and the latter at 256 v.d. If on the other hand we count the error in relative parts of a tone instead of vibration for vibration, the women sing twice as accurately as the men. It may, however, be shown that the former statement represents the more logical point of view.

This result is in harmony with the results found in Series I with reference to accuracy within the tonal range. It was there found that so long as the singer was certainly within his natural range the man could sing the two tones here considered, 128 v.d. and 256 v.d., with nearly equal accuracy, in terms of vibrations and that, therefore, he tended to sing the higher twice as accurately as the lower. The difference here discussed is therefore not peculiarly a sex difference, but distinctly a matter of psycho-physic law of voice control within the tonal range. Men and women have equal ability in pitch discrimination (reference 21 p. 44), so also in voice control they have equal ability level for level within the tonal range. The fact however remains that women's voices are pitched in a higher register than men's voices and therefore, from the musical point of view, they can sing their tones relatively more accurately.

This result is, after all what we should expect for the principal limit upon accuracy in singing is accuracy in hearing and we know that both men and women can hear a difference of, *e.g.*, 1 v.d. as easily at 256 v.d. as at 128 v.d.

The mean variation

Fig. 5 shows that the mean variation is larger for the variants than for the standards. This is because the former are more difficult. It should be noted that this difference in the mean variation is a measure of the relative difficulty of the two tones as felt

¹⁵ The following facts are significant: (1) there are fewer poor observers among the women; (2) women have smaller mean variations than men; and (3) women more nearly reproduce the intervals. It seems quite likely that in a mixed college group such as we have here, the women give more attention to vocal music than do the men, which may account for their superiority in this test.

and would also be a measure of the relative degree of accuracy in the singing of them were it not for the operation of the two motives for sharpening the variant about the middle of the series of the increments. The fact that the mean variation is unaffected by the operation of these two motives is an indication of their fairly rigid operation.

The constant error

Figs. 5 and 6 show that the singing of the standard tone is not affected by the magnitude of the increment to be sung. The constant error is small and uniform. This is due partly to the fact that the standard tone was the same in all trials and therefore tended to become more or less automatic, and partly to the fact that the standard was sung first and that therefore the difficulty in marking off the interval would tend to crop out in the variant tone.

The singing of the variant follows the law that (1) all these small increments are overestimated and that (2) this overestimation increases gradually from the largest interval (0-30) and reaches a maximum in the cases of both men and women (Fig. 5, A and B) at the 5 v.d. interval from which it gradually again diminishes.

There are probably several motives operating to produce this overestimation; the fact that the maximum falls in the increment 5 v.d. points to a relationship between the hearing and the singing of the interval. The median for the least perceptible difference in pitch for this same group of individuals falls on 3 v.d. The increment 5 v.d. in singing would therefore represent one of the smallest increments actually heard. The distribution around this would be analogous to the distribution of the records in pitch discrimination for this group.

It is probable that, as in visual perception of space, all small angles are overestimated, there is in hearing of pitch a tendency to overestimate the smallest increments perceived. If we represent the uniformly increasing series of increments of pitch difference as a sharp wedge the apparent magnitude would be represented by a wedge blunted and thickened.

The operation of such a principle has been demonstrated for hearing in the matter of localization of sound. Starch (23) found that when a correction is made for the least perceptible change in the direction of the source this correction is always overdone.

The lack of fine control of the voice to reproduce the smallest

differences that are heard is another element involved. This factor is partly due to lack of knowledge and practice in this kind of voice control. The small differences which are actually heard larger than they really are, are sung still larger on account of this general lack of control for the making of fine shadings in pitch. This overdoing of a difference may perhaps be regarded as another phase of the same principle as the overestimation of small differences in pitch in hearing. At any rate the enlarging of the small discriminated increments is without doubt much increased in the singing. These small increments are overestimated in hearing (when heard) and are again overdone in the singing; and that this enlarging is proportionate up to the threshold for pitch discrimination.

In applying these principles to the interpretation of the relative magnitude of the errors in the singing of these increments we must bear in mind that where the small differences are not heard there would be a tendency to repeat the standard in trying to sing the variant—this happens not only because the difference is not heard, but even when an effort is made to sing an imperceptible sharp theoretically known to exist there is a tendency for the voice to “fall into the groove” of the standard tone which has been sung immediately before.

On the other hand it seems reasonable to take account of the fact that in this test we are asking the observer to do something with which he is almost entirely unfamiliar. In the larger intervals he recognizes differences but overestimates and oversings them. This overestimation increases regularly from the largest interval, 0-30, to 0-5, as was above noted. At 0-3 most of the observers fail to hear the difference because the conditions of the test do not provide the immediately successive presentation which is most favorable for the discrimination of pitch differences. Therefore, at 0-3 failing to hear the second fork higher, recognizing that he has not yet reached the smallest possible interval, and knowing that the second fork is higher than the O, our observer concentrates his attention, trying harder and harder until the last interval is sung. He is in large measure freed from the factor of overestimation in hearing for he hears no difference. He will very likely tell you that the forks sound just alike, but he knows and is reminded that the second one of each pair is higher. This knowledge forms the basis of his control of the voice. Quite naturally under the circum-

stances he resorts to the tendency (noted above) to take his cue for the second tone not from the fork but from his own previous tone. He "falls into the groove", however, just long enough to get his bearings, then sharps from this point, the magnitude of the sharp being governed roughly by the subject's pitch discrimination ability. In about 8 per cent. of the individual records of 0-5 the records on the .5 v.d. are not sharp or may be slightly flat; in other words, the observers took the risk of making no sharp.

Applying these factors in the interpretation of the error in the singing of these small intervals of different magnitude, we find that, (1) the average overestimation is relatively small for the smallest increments because in many cases the difference is not heard and in singing a very small interval the voice uses its previous reproduction as the standard, sharpening from it, and (2) the overestimation of the small increment is greatest for the smallest increments perceived and gradually diminishes as the increments grow larger so that it tends to disappear on the average when the magnitude of a half-tone is reached. Therefore, our test seems to have met the conditions for measuring the minimal producible change in the pitch of the voice. The increments from 0-30 to 0-5 serve to work down the voice, to make clear to the observer what is to be done, and to center his attention for most careful control. The four smaller increments, 0-3 to 0-5 are the place where the "ability to make faint shadings" is really tested and under usual conditions the reproductions on the smallest increment, 0-5, would seem to give the best measure.

If from the records on 0-5 (algebraic C. E. or G. C. E.) we compute the magnitude of the smallest interval as *actually produced* by the individual observers and distribute these magnitudes according to their frequency, we have the curves of Fig. 12. The median value of the measures represented in Fig. 12 is 4.0 v.d. for women and 4.5 v.d. for men. There are more extremely poor observers among the men so that the average smallest intervals produced are 5.6 v.d. and 3.7 v.d. for men and women respectively.¹⁶ These median values are in harmony with the results for pitch discrimination and may be taken as measures of the ability to produce minimal changes sharp or flat in the pitch of the voice.

¹⁶ A part of this difference between men and women is to be accounted for in the fact that on the average the men flattened the O.

Dr. D. A. Anderson made a test on "minimal change in the pitch of the voice" in the Iowa Psychological Laboratory in 1909. His observers imitated the pitch of one standard fork and then sang the tone the least possible sharp or flat according as directed, making ten successive trials in each direction. There were 115 women and 65 men in the group tested. From the unpublished results of this test we learn that the average minimal producible change for men was 5.5 v.d. and for women 4.6 v.d. as against 5.6 v.d. and 3.7 v.d. in our test. In comparing these results it must however be noted that 45 of Professor Anderson's poorest observers, most of them men, made no records which entered into his averages.

Seashore (19) reports the results of some tests of "minimal producible change" given to a small group of observers. The average records for six men on five successive days are as follows; 3.4, 3.5, 3.0, 2.6, and 2.7 v.d. Evidently the factor of practice entered here. However, the average of these results, which represents the only other available data on this ability in voice control, falls on the mode of our curve (Fig. 12) for men.

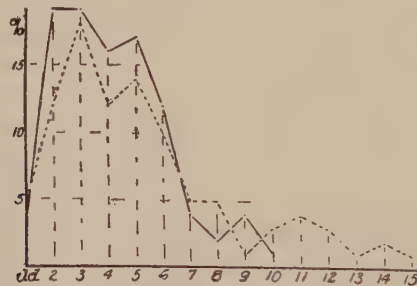


Fig. 10. Distribution of the magnitudes of the smallest interval actually produced by men and women. The method of computing the average magnitude of the smallest interval produced by each observer is illustrated in the following example: if the C. E. on O is -0.9 v.d. and on V is $+1.4$ v.d. then the produced interval would equal the difference between -0.9 v.d. and $+1.4$ v.d. plus $.5$ v.d. (the real step between the forks) = 2.8 v.d. If the C. E. on O is plus it is of course subtracted from the sum of C. E. on V and $.5$ v.d. Men dotted line; women broken line.

The average constant error (C. E.) on the standard is small and uniform, as is also the mean variation and the constant tendency for the group (G. C. E. on the standard). Accuracy in the standard is not influenced by any difference in the magnitude of the in-

crements. This is chiefly because the standard tone was sung before the variant was sounded, and partly because a sort of "rut" was formed for the singing of the repeated standard.

The researches previously reviewed contain scattered measures on this ability. Klünder (1) found that he could reproduce an organ tone of 128 v.d. with an average crude error of .47 v.d. He rejected however the records of some other observers who showed larger errors. Cameron (4) worked with seven observers and tried a number of organ tones. The records by three of these observers gave an average error of about 6.6 v.d. Berlage (2), whose three observers reproduced voice tones, does not give the pitch of the standards. The average error for the three men, singing with an interval of from 1 second to 2 seconds, is .50 v.d.¹⁷ Seashore (19) gives 1.2 v.d. as the average error of 100 trials by each of six men, on standard 100 v.d. Sokolowsky (22) with his seven professional singers finds an average error of 1 v.d. at the average pitch of 251 v.d.

The group constant error

Throughout the previous pages there have been references to the tendency of both men and women to sing sharp when reproducing a tone. The difference in the direction of this error in the standard for men and for women is so constant that, while small, it points to some motive in the character of the tone, the mode of singing, or some tendency characteristic of a given pitch level. The distribution seen in Figs. 7 and 8 shows that the sharps and the flats are not far from equal both in the number and the magnitude for men; for women the sharps predominate in both magnitude and number.

Cameron (4) noticed this tendency and called attention to it. In his experiments it appeared especially in sustained tones.¹⁸ We have not worked with sustained tones but have found the same

¹⁷ Berlage's tables are needlessly complicated by his using the signs with opposite from the usual meaning.

¹⁸ Berlage (2) did not find this tendency to sharp and was surprised, but we must remember that he worked with voice tones for standards (the richest tone possible) and our experiments seem to show that with rich standard tones the sharpening of the constant error is considerably decreased. Sokolowsky's (22) results are also negative as regards any general tendency for both sexes to sing sharp. The errors on the twenty tones sung by women, however, show an algebraic average of + 1.03 v.d., although eleven of these tones were sung flat.

tendency with reproductions of one and two seconds in length. Reference to our tables (G. C. E.) will show that almost without exception sharpening is the predominant direction of the constant errors in all six series of our experiments. The tendency to sing sharp is not materially affected by the level of the pitch so long as the tone remains within the range of the voice; it is increased by loud volume of voice, weak volume of standard, certain vowel formants such as are found in "e" and "i", and by purity of the standard tones.

The best cases

The question naturally arises, to what extent the presence of a few cases of very large error affect the averages. To cast some light on this and also to gain an idea of the performance of the best observers in the group the author made a selection of twenty-five persons of each sex. The selection was made chiefly on the basis of a small Ave. m.v. in the standard (o). The size of the Ave. C. E. of o and the Ave. m.v. for the increments, were used as secondary criteria. There are some records, for example N. 9, which from the standpoint of the constant errors alone are very near the ideal curves, but because of rather large mean variations must be omitted from these selected groups. The selection of women was as follows: Nos. 1, 3, 13, 15, 21, 55, 61, 63, 77, 85, 93, 97, 105, 107, 113, 117, 125, 153, 159, 169, 177, 181, 183, 201, and 209. The men's records chosen were: Nos. 6, 8, 10, 12, 16, 28, 50, 62, 68, 72, 82, 88, 102, 106, 110, 114, 120, 126, 128, 144, 146, 148, 154, 156, and 164.

The separate tabulation of these fifty supposedly best cases reveals the presence of the same general tendencies in these selected groups as have been noted in the large groups, with the difference that they are not so pronounced and that here the men in a relative comparison make a better showing than the women, in that their overestimation especially of the smaller pitch increments is less. Therefore blame for the large errors (overestimation of intervals) can hardly be shifted to a few individuals as indeed we might have shown by referring to Figs. 7 and 8 which demonstrate that the distribution of the errors forms fairly normal frequency curves.

Correlation of singing with pitch discrimination

Pitch discrimination records are available for eighty-two of the men, and one hundred and four of the women who acted as observers

in our tests. The well-known formula of the Pearson "Product-Moments" was employed and resulted in the following correlations:

						r	P.E.r
Men:	Size of ave.	smallest interval	produced with	Pitch.	Disc.	+.21	.072
	" " "	m.v. on O.	"	"	"	+.04	.074
	" " "	C.E. on O.	"	"	"	+.08	.074
	" " "	m.v. on V.	"	"	"	+.33	.066
	" " "	C.E. on V.	"	"	"	+.15	.073
Women:	" " "	smallest interval	produced	"	"	-.11	.065
	" " "	m.v. on O.	"	"	"	+.27	.061
	" " "	C.E. on O.	"	"	"	+.11	.065
	" " "	m.v. on V.	"	"	"	+.51	.048
	" " "	C.E. on V.	"	"	"	-.07	.065

It will be recalled that in order to be satisfactory a coefficient "should be perhaps three to five times as large" as its probable error. This rule liberally applied to our results leaves us the coefficients +.33 and +.51 both of unquestionable reliability. These coefficients represent the correlation between pitch discrimination and the average mean variation in singing the intervals, for men and women respectively.

The Test of 1910

A series of musical tests, given by the writer in the Iowa Psychological Laboratory, during November and December of 1910, included one on Accuracy in Reproducing Tones. There were ninety men and one hundred and seven women, members of the elementary psychology classes who took this test.

The apparatus besides the tonoscope consisted of five large forks with pitches as follows: 128, 256, 320, 384, and 512 v.d. The experimenter instructed the observer to take the 256 v.d. fork, strike it gently, bring it to his ear, listen carefully, and then to reproduce the same pitch. This he repeated with fork 256 v.d. Then taking the 320 v.d. fork he proceeded as described. The last four forks were gone over five times in this manner, which gives ten trials on each tone, forty trials in all. The test is thus very simple, the reproduction of four tones (two successive trials on each) which are at the same time natural musical intervals: major third, fifth and octave. No restrictions were placed upon the observer in the matter of humming or singing with the standards. As the fork was in his hand he sang with it or after it as seemed best to him. About one-half the observers preferred to take the fork away from the ear before beginning to sing. The men sang the tones one octave below

the pitch of the fork. When it was difficult for them to commence doing this, the 128 v.d. standard was used for orientation.

The results for this series of 8,000 reproductions are given in Table IX. The notation is the same as in Table VIII. The test of 1910 was complicated by the factor of natural musical intervals, it was also considerably shorter and simpler than the one of 1913 but in comparing it with the latter we find the results in practical agreement on some points.

(1). There is a uniform tendency for the majority of observers to sing sharp. Here again the tendency appears to be greater for women than for men, the G. C. E. for men being $+.26$ v.d., for

TABLE IX. *Accuracy of singing: test of 1910*

90 men	128 v.d.	160 v.d.	192 v.d.	256 v.d.
Ave. m.v.	1.42	1.36	1.43	1.79
% + C. E.	47	63	51	63
% - C. E.	53	37	49	37
Av. C. C.	1.62	1.62	1.72	2.70
G. C. E.	$+.26$	$+.65$	$-.06$	$+.06$
107				
women	256 v.d.	320 v.d.	384 v.d.	512 v.d.
Ave. m.v.	1.89	2.15	2.07	2.83
% + C. E.	81	90	84	86
% - C. E.	19	10	16	14
Av. C. E.	2.59	3.91	3.47	5.90
G. C. E.	$+2.39$	$+3.36$	$+3.11$	$+4.76$

women $+2.39$ v.d., a difference of 2.13 v.d. as contrasted with 1.30 v.d. in the previous measurements. In the test of 1910, as mentioned, the men and women used the same forks, the men singing the standards one octave low. Therefore the tendency for men to sing less sharp than women in the 1913 experiments can hardly be attributed to a timbre or sound volume difference between the sets of forks. The men are must more evenly divided between the sharpening and flattening tendencies than the women, for example on 256 v.d. the one tone which both sexes had in common, the percentages in favor of sharpening are 63 and 86 for men and women respectively. (2) The average constant error (arithmetic) on 128 v.d. is slightly larger in 1910, 1.62 v.d. as against 1.54 v.d. The mean variation for 128 v.d. are 1.42 v.d. (1910) and 1.54 v.d. These differences are rather slight. (3) Men and women sing their one common tone (256 v.d.)

with equal accuracy: m.v. 1.79 v.d., Av. C. E., 2.70 v.d. (men) to m.v., 1.89 v.d., Av. C. E. 2.59 (women). It would seem from a comparison of available norms for voice range in the sexes (Helmholtz (9) and Zahm (27) that 256 v.d. should be about as high for men as it is low for women, and that it is well within the average range of both. We have here therefore a confirmation of our previous conclusion, *i.e.*, that men and women sing with equal accuracy vibration for vibration. However the errors in this case under consideration (1910) are much larger than the results of Series VI would lead us to expect. This is true of all the tones sung by the women and renders them incomparable with the previous results.

Recommendations toward a standard test

The recommendations which follow must be considered as preliminary and as applying simply to the two measures of singing ability considered throughout this study, *i.e.*, the ability of the voice to reproduce pitch, and the ability to produce voluntarily small changes sharp or flat in the pitch of the voice.

1. The two factors may be tested together with advantage. They are closely related phases of the same thing. Neither of them should be taken in combination with such factors as accuracy of tone memory, or judgment for musical intervals.

2. Use a graded series of standard tones similar to that commonly employed in testing for pitch discrimination. Such a series has obvious advantages over the use of a single standard; (1) If several observations are to be made at a single sitting the effects of practice are not so great. (2) The small pitch intervals make clear to the observer what he is expected to do with his voice. (3) The variety of standards (and hence degrees of difficulty) reduce monotony and fatigue. A graded series furthermore has advantage over any other series: (1) it keeps the test comparatively free from complication with the singing of musical intervals, and (2) when the standards represent small steps of pitch difference the observer discriminates more carefully and is not so likely to be satisfied with a mere approximation.

3. Use tuning forks for standards. They are very easily manipulated, are not subject to certain sources of error commonly met in the control of reeds, pipes and strings, and are readily arranged

into a graded series as recommended above. Any disadvantage, if indeed it may be so called, from the standpoint of the purity of the fork tone seems more than compensated for in having a definable quality and a standard on which all observers are equally unpracticed.

4. Begin with the largest pitch increments and proceed to the smallest and then in reverse order back to the largest. This will economize effort, provide the best practice, and help to control the attention. For general testing ten intervals representing as many degrees of difficulty, ranging from 0-30 to 0-5 are not too many. For extensive testing of one observer or in working with highly practiced observers the increments which are distinctly above the threshold for pitch discrimination may be omitted.

5. Give the tones in pairs, presenting the variant tone immediately after the reproduction of the standard, thus securing a rapid adjustment which favors discrimination in the kinaesthetic sensations from the larynx. As an alternative procedure the two tones might be presented in immediate succession as in the pitch discrimination tests, the observer carrying the standard in mind while listening to the variant, and then singing them in quick succession.

6. Control conditions: (1) The forks should be presented before resonators which are some distance from the observer and care must be exercised to present them with uniform intensity. (2) The observer should use a medium volume of voice in singing the tones. (3) The experimenter should select the vowel to be sung and insist on a good quality. (4) If time intervals are used between standards and reproductions they should be short, not longer than two seconds at most. (5) Time intervals should be introduced between pairs of tones. These should be at least 2 seconds in length. Longer intervals would doubtless be better as the voice could the more easily be kept out of a "rut" in reproducing the standard. (6) Secure effort on the part of the observer who is too easily satisfied with his own performance.

Our test is one of motor control. As a musical test it bears the same relation to the motor side as pitch discrimination does to the sensory side. In fact it is in a practical way for the motor pitch discrimination of the singer, and as far as singing is concerned it is more important than simple sensory pitch discrimination.

SUMMARY OF CONCLUSIONS

Among others the following general conclusions may be gleaned from the foregoing experiments.

1. The human voice is about equally accurate, in terms of vibration, at all points well within its range; therefore, the high tones are sung relatively (per cent.) more exactly than those which are low.

2. A strong standard tone (especially with low forks) is reproduced as decidedly lower than a weak standard.

3. The voice can most easily reproduce pitch for those standard tones which have a rich timbre, such as the organ tone.

4. Measured in terms of average error the voice is less accurate when its volume is large.

5. Vowel quality affects the accuracy of vocal reproduction of tones. The "a" (as i in machine) is reproduced the highest, "o" the lowest, and "u" occupies a middle position.

6. Men and women sing in their representative ranges with equal accuracy vibration of error.

7. Women show better relative voice control than men, if judged on the basis of their mean variation.

8. With women there is a general tendency to sing sharp. Men are about equally divided in this regard, sharpening however being slightly more frequent.

9. The average error of the voice in reproducing a tone given by a fork is 1.5 v.d. for men at range 128 v.d., and 1.5 v.d. for women at 256 v.d. in a representative group of students.

10. A small perceptible pitch difference between two tones is overestimated in the singing.

11. The average minimal producible change of the voice for men at 128 v.d. is about 5.5 v.d., and for women at 256 v.d. it is 3.5 v.d.

REFERENCES

1. Anderson, D. A. The Effect of Intensity on the Apparent Pitch of Tones in the Middle Range. (In this Volume.)
2. Blake, The Use of the Membrana Tympani as a Phonauto-graph and Logograph. *Arch. of Opthal. and Otol.*, 1876, V.
3. Berlage, F. Der Einfluss von Artikulation und Gehör beim Nachsingen von Stimmklängen. *Psychol. Stud.*, 1910, VI, 39-140.
4. Cameron, E. H. Tonal Reactions. *Psychol. Rev., Monog. Suppl.*, 1907, VIII, 227-300.
5. Grutzner, *Die Stimme*. 1907.

6. Guttman, A. Zur Psychophysik des Gesanges. *Zsch. f. Psychol. u. Physiol. d. Sinnes.*, 1913, LXIII, 161-176.
7. Gutzmann, H. A. K. *Physiologie der Stimme und Sprache*. Braunschweig, 1909.
8. Hancock, C. L. The Effect of Intensity on the Apparent Pitch of Tones in the Lower Range. (In this Volume.)
9. Helmholtz, H. *Sensations of Tone*. Tr. by Ellis 1895.
10. Hensen. Ein einfaches Verfahren zur Beobachtung der Tonhöhe eines gesungen Tons. *Arch. f. Anat. u. Physiol. (Physiol. Abth.)*, 1879, pp. 155 ff.
11. Klünder, A. Ein Versuch die Fehler zu bestimmen welche der Kehlopf beim Halten eines Tones macht. Inaug. Dissertation, Marburg, 1872.
12. Klünder, Ad. Ueber die Genauigkeit der Stimme. *Arch. f. Anat. u. Psychol.*, Du Bois Reymond, (Physiol. Abth.) 1879, pp. 119, ff.
13. Luft, E. Ueber die Unterschiedsempfindlichkeit für Tonhöhen, *Phil. Stud.* 1888, IV, 511-540.
14. Marbe, K. Über die Verwendung russender Flammen in der Psychologie und deren Grengebieten, *Zsch. f. Psychol.*, 1908, XLVIX, 206-217.
15. Meyer, Max. Ueber die Unterschiedsempfindlichkeit für Tonhöhen, *Ztsch. f. Psychol.*, 1898, XVI, 352-372.
16. Preyer,
17. Scott. *Inscription automatique des sons de l'air au moyen d'une oreille artificielle*. 1861.
18. Scripture, E. W. *Researches in Experimental Phonetics*. Carnegie Institution, 1906.
19. Seashore, C. E. and Jenner, E. A. Training the Voice by the Aid of the Eye in Singing. *J. of Educ. Psychol.*, 1910, I, 311-320.
20. Seashore, C. E. A Voice Tonoscope. *Iowa Stud. in Psychol.*, 1902, III, 18 ff.
21. Seashore, C. E. The Measurement of Pitch Discrimination: A Preliminary Report. *Psychol. Monog.*, XIII, 21-60.
22. Sokolowsky, R. Ueber die Genauigkeit des Nachsingens von Tönen bei Berufssängern. *Beitr. z Anat., Physiol., Path, usw. von Passow u. Schaefer.*, 1911, V.
23. Starch, D. Perimetry of the Localization of Sound. (*Iowa Stud. in Psychol.*, V, 1-55) *Psychol. Rev. Monog.*, IX.
24. Stern, H. Gesangphysiologie und Gesangpädagogik ihren Beziehungen zur Frage der Muskelempfindungen und der beim Singen am Schädel und am Thorax fühlbaren Vibrationen. *Monat. f. Ohrenhk.*, 1912, XLVI, 337-352.
25. Stumpf, C. *Tonpsychologie*. 1883.
26. Vance, T. F. Pitch Discrimination within the Tonal Range. (In this Volume.)
27. Zahm, J. A. *Sound and Music*. Chicago; McClurg 1892.

